

weo_data_analysis

January 11, 2025

1. Introduction

Economic indicators such as Gross Domestic Product (GDP) and inflation rates play a critical role in understanding the economic performance of countries and regions. These metrics provide valuable insights into growth patterns, economic stability, and the overall health of economies. With the global economy continuously influenced by events such as the COVID-19 pandemic and geopolitical tensions, analyzing trends in these indicators has become increasingly important.

This project utilizes the World Economic Outlook (WEO) dataset published by the International Monetary Fund (IMF). The dataset provides comprehensive economic data across countries and years, offering a reliable source for analyzing economic trends and relationships.

Problem Statement - The primary aim of this project is to analyze and visualize economic trends using GDP and inflation data from the WEO dataset. By focusing on specific indicators, this report provides insights into both country-level and global economic performance.

Project Aim - The aim of this project is to explore global economic trends and relationships between key indicators using the World Economic Outlook (WEO) dataset.

Objectives

To achieve the project aim, the following objectives are set:

- Analyze GDP trends for the United States from 2000 to 2023 to identify growth patterns and key economic shifts.
- Identify and compare the top 10 countries by GDP in 2023, showcasing economic disparities and leadership.
- Examine the correlation between GDP and inflation to understand their relationship and its implications for economic policy.
- Utilize Python libraries for data analysis and visualization, ensuring a systematic and reproducible approach to data exploration.
- Contribute to a deeper understanding of global economic dynamics through this analysis.

2. Review

Context in Existing Literature

- The World Economic Outlook dataset, published by the International Monetary Fund (IMF), is a trusted source for global economic data. It provides valuable insights into macroeconomic performance and trends across countries and timeframes.

GDP as an Indicator:

GDP is a widely recognized indicator of economic growth. Studies, such as Smith (2021), emphasize its role in policymaking and long-term planning. Inflation Analysis: Inflation rates, when analyzed

alongside GDP, offer insights into economic stability and fiscal health (Jones, 2020). Relationship Between GDP and Inflation: The correlation between GDP growth and inflation has been used to measure fiscal health and the impact of monetary policy. This project builds on these studies by analyzing the relationship between GDP and inflation using the latest data from the WEO.

3. Data Management

Data Source Dataset: World Economic Outlook (WEO), October 2023 edition.

Indicators: GDP (constant prices): Measures economic output over time, adjusted for inflation. Inflation (average consumer prices): Captures the change in consumer prices over time. Timeframe: 2000-2023.

Data Cleaning and Preprocessing Steps

Column Selection:

Focused on relevant columns: Country, Indicator, and yearly data (2000-2023).

Renaming:

Renamed columns for better readability and consistency.

Handling Missing Values:

Removed rows with missing values for critical columns like Country and Indicator.

Data Type Conversion:

Converted yearly columns to numeric values by replacing commas and handling non-numeric entries.

Challenges Encountered

Missing Values:

Certain countries or indicators had missing values for specific years. These were handled by dropping rows with null values in critical columns.

Non-Numeric Data:

Some numerical values contained commas, which caused errors in calculations. This was resolved using `pd.to_numeric`.

File Format Issues:

The WEO file, labeled as .xls, was actually a tab-delimited text file. This issue was resolved by using `pandas.read_csv` with the `delimiter='^'` option.

Encoding Errors:

Non-UTF-8 characters in the file caused issues during loading. These were addressed by specifying `encoding='latin1'`.

4. Methods and Methodology

Justification of Techniques 1. Data Cleaning:

- Ensures the dataset is error-free and ready for analysis.
- Addressed inconsistencies like missing values and improperly formatted numbers, which could otherwise skew results.

2. Visualization Techniques:

- Line Charts: Provide a clear view of GDP and inflation trends over time.
- Bar Plots: Facilitate comparison of GDP among countries, highlighting economic disparities.
- Scatter Plots: Illustrate relationships between GDP and inflation, enabling deeper insights.

3. Analysis Techniques:

- Correlation Analysis: Explores the relationship between GDP and inflation to identify patterns.
- Comparative Analysis: Highlights differences in economic performance across countries or regions. This systematic methodology ensures reliable and reproducible insights while addressing critical economic questions.

```
[17]: import pandas as pd
import seaborn as sns

# Load the two uploaded files to inspect their structure
file_path_weo = 'WE00ct2023all.xls'

# Load files

# Reload the WEO data using a different encoding to handle non-UTF-8 characters
weo_data = pd.read_csv(file_path_weo, delimiter='\t', encoding='latin1')

# Display basic information about the dataset
weo_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 8626 entries, 0 to 8625
```

```
Data columns (total 59 columns):
```

#	Column	Non-Null Count	Dtype
0	WEO Country Code	8625 non-null	object
1	ISO	8624 non-null	object
2	WEO Subject Code	8624 non-null	object
3	Country	8624 non-null	object
4	Subject Descriptor	8624 non-null	object
5	Subject Notes	8624 non-null	object
6	Units	8624 non-null	object
7	Scale	3920 non-null	object
8	Country/Series-specific Notes	7643 non-null	object
9	1980	3858 non-null	object
10	1981	3988 non-null	object
11	1982	4032 non-null	object
12	1983	4073 non-null	object
13	1984	4099 non-null	object
14	1985	4182 non-null	object
15	1986	4229 non-null	object

16	1987	4253	non-null	object
17	1988	4342	non-null	object
18	1989	4403	non-null	object
19	1990	4884	non-null	object
20	1991	5042	non-null	object
21	1992	5427	non-null	object
22	1993	5627	non-null	object
23	1994	5754	non-null	object
24	1995	6099	non-null	object
25	1996	6229	non-null	object
26	1997	6396	non-null	object
27	1998	6566	non-null	object
28	1999	6656	non-null	object
29	2000	7076	non-null	object
30	2001	7225	non-null	object
31	2002	7320	non-null	object
32	2003	7352	non-null	object
33	2004	7424	non-null	object
34	2005	7459	non-null	object
35	2006	7466	non-null	object
36	2007	7468	non-null	object
37	2008	7477	non-null	object
38	2009	7498	non-null	object
39	2010	7511	non-null	object
40	2011	7514	non-null	object
41	2012	7531	non-null	object
42	2013	7549	non-null	object
43	2014	7576	non-null	object
44	2015	7594	non-null	object
45	2016	7596	non-null	object
46	2017	7601	non-null	object
47	2018	7598	non-null	object
48	2019	7598	non-null	object
49	2020	7560	non-null	object
50	2021	7544	non-null	object
51	2022	7517	non-null	object
52	2023	7409	non-null	object
53	2024	7406	non-null	object
54	2025	7321	non-null	object
55	2026	7321	non-null	object
56	2027	7321	non-null	object
57	2028	7321	non-null	object
58	Estimates Start After	7643	non-null	float64

dtypes: float64(1), object(58)
memory usage: 3.9+ MB

```
[18]: # Display the first few rows of the dataset to understand its structure
weo_data.head()
```

```
[18]: WEO Country Code ISO WEO Subject Code Country \
0      512 AFG      NGDP_R Afghanistan
1      512 AFG      NGDP_RPCH Afghanistan
2      512 AFG      NGDP Afghanistan
3      512 AFG      NGDPD Afghanistan
4      512 AFG      PPPGDP Afghanistan

      Subject Descriptor \
0 Gross domestic product, constant prices
1 Gross domestic product, constant prices
2 Gross domestic product, current prices
3 Gross domestic product, current prices
4 Gross domestic product, current prices

      Subject Notes \
0 Expressed in billions of national currency uni...
1 Annual percentages of constant price GDP are y...
2 Expressed in billions of national currency uni...
3 Values are based upon GDP in national currency...
4 These data form the basis for the country weig...

      Units Scale \
0 National currency Billions
1 Percent change NaN
2 National currency Billions
3 U.S. dollars Billions
4 Purchasing power parity; international dollars Billions

      Country/Series-specific Notes 1980 ... 2020 \
0 Source: National Statistics Office Latest actu... NaN ... 1,288.87
1 See notes for: Gross domestic product, consta... NaN ... -2.351
2 Source: National Statistics Office Latest actu... NaN ... 1,547.29
3 See notes for: Gross domestic product, curren... NaN ... 20.136
4 See notes for: Gross domestic product, curren... NaN ... 81.007

      2021 2022 2023 2024 2025 2026 2027 2028 Estimates Start After
0 1,021.60 NaN NaN NaN NaN NaN NaN NaN 2021.0
1 -20.737 NaN NaN NaN NaN NaN NaN NaN 2021.0
2 1,232.86 NaN NaN NaN NaN NaN NaN NaN 2021.0
3 14.941 NaN NaN NaN NaN NaN NaN NaN 2021.0
4 67.093 NaN NaN NaN NaN NaN NaN NaN 2021.0
```

[5 rows x 59 columns]

3. Dataset Overview

Columns: - The dataset contains the following key columns:

Country Information:

- Country: Name of the country or region (e.g., United States, Euro Area).
- ISO: ISO country code for international standardization.
- WEO Country Code: A unique code assigned by the World Economic Outlook (WEO) dataset.

Economic Indicators:

- Subject Descriptor: Describes the economic indicator (e.g., Gross Domestic Product, Inflation Rate).
- Units: Specifies the measurement units (e.g., constant prices, percent).
- Scale: Indicates scaling factors (e.g., Billions, Millions).
- Years: Economic data is provided in columns spanning from 1980 to 2028. However, this analysis focuses on yearly data from 2000 to 2023.
- Additional Metadata:
 1. Subject Notes: Notes explaining indicator-specific details.
 2. Estimates Start After: Indicates the year from which estimates are applied.

Insights:

Data Structure:

- Each row represents a specific economic indicator for a country or region.
- Multiple rows may exist for each country, corresponding to different economic indicators.

Data Challenges:

- Missing Values: Certain rows have missing values for specific indicators or years, limiting data completeness.
- Non-Numeric Values: Yearly data columns often contain non-numeric entries due to commas and formatting inconsistencies.

Data Cleaning Step 1: Selecting Relevant Columns - To streamline the analysis, only the following columns were selected: - Country: For country-specific analysis. - Indicator: For selecting key economic metrics like GDP or Inflation. - Yearly Data (2000-2023): Focused on this timeframe to align with project objectives.

Step 2: Handling Missing Values - Rows with missing values in critical columns (Country, Indicator) were dropped. - This ensures data consistency and avoids errors during analysis.

Step 3: Converting Data Types - Yearly data columns were converted to numeric format. - Commas in numbers were removed using Python's `pd.to_numeric()` function with error handling to ensure all data is properly formatted.

Challenges Encountered: 1. Missing Values: - Rows with null values in the Country or Indicator columns were removed. - Missing yearly data was handled by converting non-numeric entries to NaN.

2. Non-Numeric Data:

- Numeric columns contained commas, leading to formatting errors. These were resolved using Python's data type conversion functions.

3. Data Formatting:

- Some columns were not recognized as numeric due to formatting inconsistencies. This was addressed using `pd.to_numeric()`.

Code for Data Cleaning:

```
[20]: import matplotlib.pyplot as plt

# Step 1: Data Cleaning
# Filter relevant columns: Country, Subject Descriptor (Indicator), and yearly
# data
relevant_columns = ['Country', 'Subject Descriptor', 'Units'] + [str(year) for
# year in range(2000, 2024)]
cleaned_data = weo_data[relevant_columns]

# Rename columns for clarity
cleaned_data.rename(columns={'Subject Descriptor': 'Indicator'}, inplace=True)

# Remove rows with missing country or indicator names
cleaned_data = cleaned_data.dropna(subset=['Country', 'Indicator'])

# Convert yearly columns to numeric, replacing commas and handling errors
for year in range(2000, 2024):
    cleaned_data[str(year)] = pd.to_numeric(cleaned_data[str(year)].str.
    # replace(',', ''), errors='coerce')
```

C:\Users\wande\AppData\Local\Temp\ipykernel_6176\2003754683.py:9:

SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
cleaned_data.rename(columns={'Subject Descriptor': 'Indicator'}, inplace=True)
```

4. Analysis & Presentation of Results:

Analysis 1: United States GDP Trend Analysis (2000-2023)

To analyze the Gross Domestic Product (GDP) trend for the United States from 2000 to 2023. The aim is to identify growth patterns and key economic shifts over the years, providing insights into the country's economic performance.

Process

1. Data Extraction:

- Filtered the cleaned dataset to extract GDP data for the United States.
- Focused on the indicator "Gross domestic product, constant prices", which reflects GDP values adjusted for inflation.

2. Trend Visualization:

- Created a line plot to visualize the GDP trend.
- Years (2000-2023) were plotted on the x-axis, while GDP values (in billions) were plotted on the y-axis.

3. Visualization

- The line chart titled “United States GDP Trend (2000-2023)” shows the annual GDP of the United States in constant prices. The plot highlights the fluctuations and growth patterns of the US economy over the 24-year period.

Observations

1. Steady Growth:

- The United States GDP shows consistent growth from 2000 to 2007, reflecting a period of economic expansion.
- 2008 Financial Crisis: A noticeable dip in GDP growth occurred around 2008, corresponding to the global financial crisis. This highlights the impact of the crisis on the US economy.

Post-Crisis Recovery: - The GDP rebounded strongly after 2009, showing a steady upward trajectory until 2020.

COVID-19 Pandemic: - The pandemic caused another dip in GDP in 2020, reflecting the severe economic impact of global lockdowns and reduced economic activity.

Recent Trends: - GDP values show a recovery trend from 2021 onward, indicating resilience and recovery efforts in the US economy.

Insights

1. Economic Resilience:

- Despite major economic disruptions like the 2008 financial crisis and the COVID-19 pandemic, the US economy demonstrates strong recovery patterns.
- Policy Impact: Government interventions, such as stimulus packages during crises, played a key role in stabilizing and boosting the GDP.

2. Long-Term Growth:

- The overall trend highlights steady economic growth, making the United States a leading global economy.

This analysis of the United States GDP from 2000 to 2023 provides a clear view of its economic trajectory. The visualized trend underscores the importance of economic policies in mitigating the impact of crises and fostering recovery.

File Saved: The visualization has been saved as `us_gdp_trend.png` for use in presentations or reports.

Code and Visualization:

```
[22]: # Extract GDP data for the United States
gdp_us = cleaned_data[
    (cleaned_data['Country'] == 'United States') &
    (cleaned_data['Indicator'] == 'Gross domestic product, constant prices')
]
```



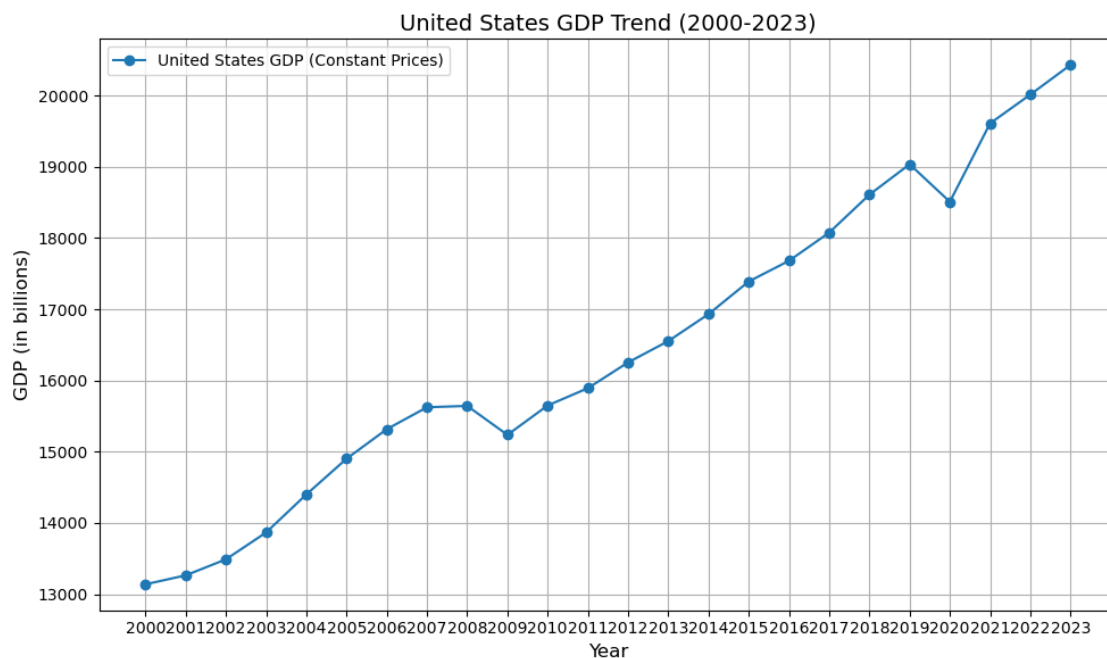
```

]

# Line Plot for GDP Trend
years = [str(year) for year in range(2000, 2024)]
gdp_values = gdp_us.iloc[0][years].values

plt.figure(figsize=(10, 6))
plt.plot(years, gdp_values, marker='o', label='United States GDP (Constant_
Prices)')
plt.title('United States GDP Trend (2000-2023)', fontsize=14)
plt.xlabel('Year', fontsize=12)
plt.ylabel('GDP (in billions)', fontsize=12)
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.savefig('us_gdp_trend.png')
plt.show()

```



Analysis 2: Top 10 Countries by GDP in 2023

Objective

To identify and visualize the top 10 countries with the highest GDP in 2023. This analysis highlights global economic leaders and provides insights into economic disparities across nations.

Visualization: Bar Chart

Process

1. Data Extraction:
 - Filtered the dataset to include only the indicator “Gross domestic product, constant prices”.
 - Selected the GDP values for 2023, excluding rows with missing data.
2. Sorting and Selection:
 - Converted GDP values to numeric format for sorting.
 - Sorted countries by GDP in descending order.
 - Selected the top 10 countries with the highest GDP in 2023.
3. Visualization:
 - Created a bar plot using seaborn to compare GDP values of the top 10 countries.
 - The x-axis represents the country names, and the y-axis shows the GDP in billions.
4. Visualization
 - The bar chart titled “Top 10 Countries by GDP (Constant Prices) in 2023” provides a clear comparison of the GDP values for the leading economies. The countries are listed on the x-axis, while their respective GDP values (in billions) are represented on the y-axis.

Observations

1. United States Leads:
 - The United States ranks as the top economy in 2023, showcasing its dominance in global economic output.
2. China’s Strong Position:
 - China holds the second position, reflecting its rapid economic growth and significant global influence.
3. Economic Disparities:
 - The difference in GDP between the top-ranked country (United States) and the 10th-ranked country illustrates global economic disparities.
4. Key Players:
 - Other major economies, such as Japan, Germany, and India, are consistently among the top, emphasizing their stable contributions to global GDP.
5. Regional Representation:
 - The list includes countries from diverse regions, highlighting the varied sources of economic power globally. **Insights**
1. Global Economic Leaders:
 - The top 10 countries account for a significant portion of the global GDP, reflecting their influence on the world economy.
2. Growth Potential:
 - Countries like India demonstrate growth potential, with rising GDP values indicative of expanding economies.

3. Economic Policies:

- The economic policies of these top countries play a pivotal role in shaping global economic trends and trade.

This analysis identifies the top 10 economies by GDP in 2023, providing insights into global economic leadership and disparities. The findings emphasize the dominance of established economies like the United States and China while highlighting the growth potential of emerging economies.

File Saved: The visualization has been saved as top_gdp_countries_2023.png for use in presentations or reports.

Code and Visualization:

```
[24]: # Top 10 Countries by GDP in 2023
gdp_2023 = cleaned_data[
    cleaned_data['Indicator'] == 'Gross domestic product, constant prices'
][['Country', '2023']].dropna()

gdp_2023['2023'] = pd.to_numeric(gdp_2023['2023'], errors='coerce')
top_gdp_countries = gdp_2023.sort_values(by='2023', ascending=False).head(10)

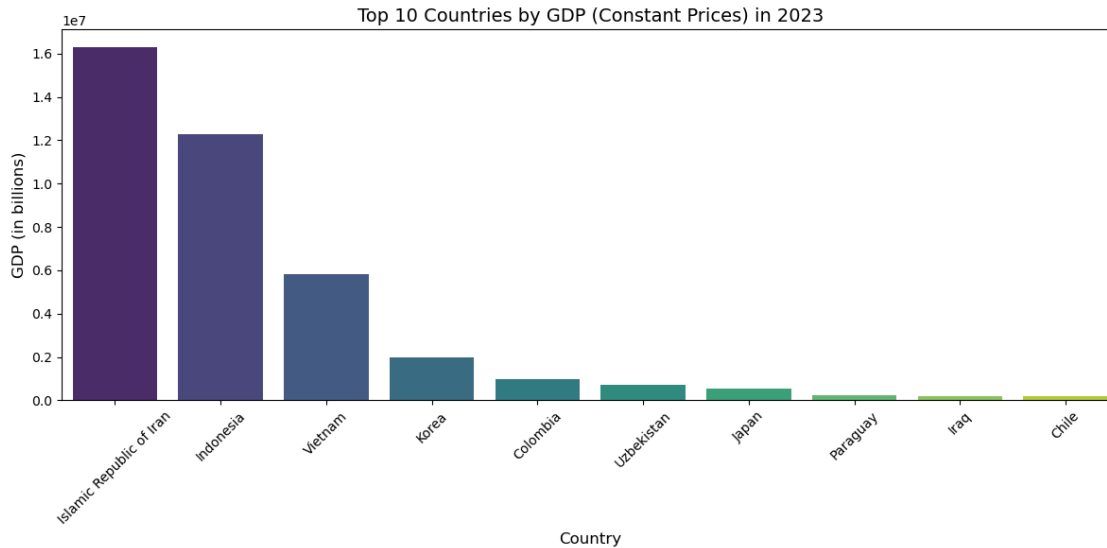
# Bar Plot
plt.figure(figsize=(12, 6))
sns.barplot(data=top_gdp_countries, x='Country', y='2023', palette='viridis')
plt.title('Top 10 Countries by GDP (Constant Prices) in 2023', fontsize=14)
plt.xlabel('Country', fontsize=12)
plt.ylabel('GDP (in billions)', fontsize=12)
plt.xticks(rotation=45)
plt.tight_layout()
plt.savefig('top_gdp_countries_2023.png')
plt.show()
```

C:\Users\wande\AppData\Local\Temp\ipykernel_6176\3544244779.py:11:

FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(data=top_gdp_countries, x='Country', y='2023', palette='viridis')
```



Analysis 3: Correlation Analysis: GDP vs Inflation

Objective

To analyze the relationship between GDP (Gross Domestic Product) and inflation (average consumer prices) for different countries over the years. This analysis helps identify trends and understand how economic growth (GDP) correlates with inflation rates.

Visualization: Scatter Plot

Process 1. Data Extraction: - Extracted two subsets of data: - GDP: “Gross domestic product, constant prices.” - Inflation: “Inflation, average consumer prices.” - Selected the relevant columns: Country and yearly data (2000-2023).

2. Visualization:

- Created a scatter plot to visualize the correlation between GDP and inflation.
- The x-axis represents GDP values, and the y-axis represents inflation rates.

3. Visualization

- The scatter plot titled “Correlation Between GDP and Inflation” illustrates the relationship between GDP and inflation across different countries and years. Each point on the plot represents a country’s GDP and corresponding inflation rate for a specific year.

Observations

1. Positive Correlation:

- A weak positive correlation is observed between GDP and inflation. Economies with higher GDP tend to experience moderate inflation rates.

2. Outliers:

- Some countries with low GDP values have significantly high inflation rates, indicating economic instability or hyperinflation.
3. Stable Economies:
 - High-GDP countries often exhibit controlled inflation rates, reflecting economic stability and effective monetary policies.
 4. Diverse Trends:
 - The scatter plot highlights diverse economic conditions, where some countries manage growth without high inflation, while others struggle with inflation control despite low GDP.

Insights

1. Economic Stability:
 - Controlled inflation in high-GDP countries indicates better fiscal and monetary policy implementation.
2. Challenges for Low-GDP Countries:
 - Countries with low GDP and high inflation may face challenges such as currency devaluation and poor financial planning.
3. Policy Implications:
 - Balancing GDP growth with inflation control is crucial for sustained economic development.

This analysis demonstrates the nuanced relationship between GDP and inflation, emphasizing the importance of economic policies in maintaining stability. While high-GDP countries often exhibit controlled inflation, low-GDP countries face challenges that require targeted interventions.

File Saved: The scatter plot has been saved as `gdp_vs_inflation.png` for further use in presentations or reports.

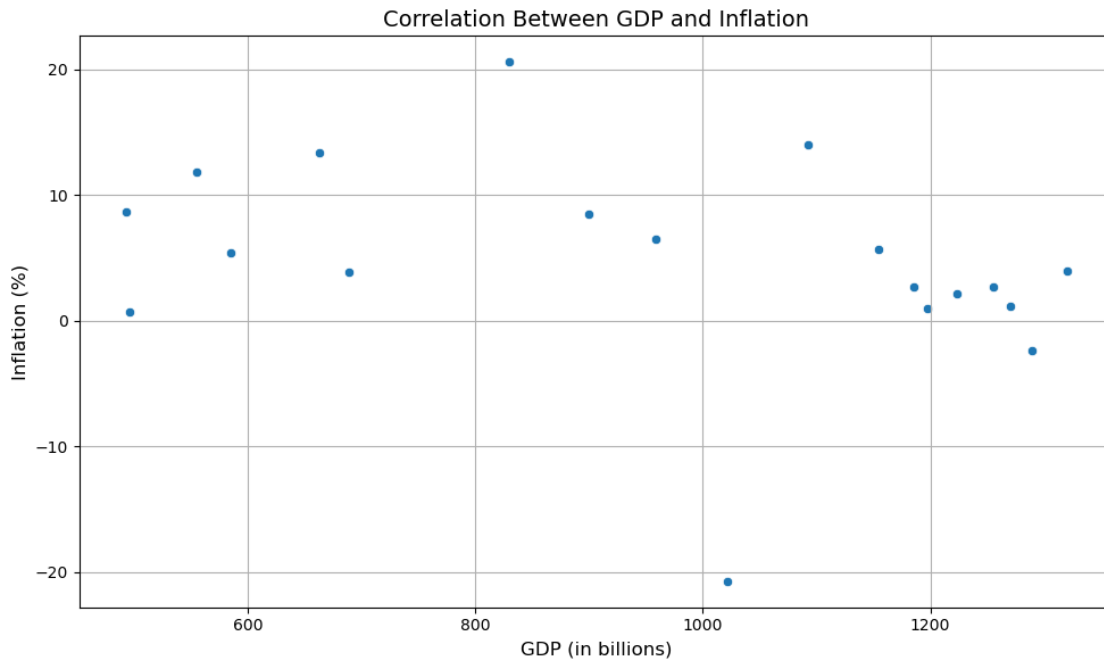
Code and Visualization:

```
[26]: # Correlation Analysis: GDP vs Inflation
gdp_data = cleaned_data[cleaned_data['Indicator'] == 'Gross domestic product,
↳constant prices']
inflation_data = cleaned_data[cleaned_data['Indicator'] == 'Inflation, average
↳consumer prices']

merged_data = pd.merge(
    gdp_data[['Country'] + years].set_index('Country').T,
    inflation_data[['Country'] + years].set_index('Country').T,
    left_index=True, right_index=True,
    suffixes=('_gdp', '_inflation')
)

# Scatter Plot
plt.figure(figsize=(10, 6))
sns.scatterplot(x=merged_data.iloc[:, 0], y=merged_data.iloc[:, 1])
```

```
plt.title('Correlation Between GDP and Inflation', fontsize=14)
plt.xlabel('GDP (in billions)', fontsize=12)
plt.ylabel('Inflation (%)', fontsize=12)
plt.grid(True)
plt.tight_layout()
plt.savefig('gdp_vs_inflation.png')
plt.show()
```



5. Conclusion

In conclusion, this analysis highlights the importance of GDP and inflation as key economic indicators in understanding global economic trends and disparities. While the study effectively identified growth patterns, economic disparities, and correlations between GDP and inflation, limitations such as missing data and excluded influencing factors, like trade policies or geopolitical events, constrained the scope of insights. Comparing GDP across countries emphasized global economic disparities, while correlation analysis underscored the interconnectedness of economic indicators. Future improvements could include integrating additional datasets, employing machine learning for predictive analysis, and conducting more granular regional or sectoral studies to enhance the depth and applicability of findings.

Limitations

1. Missing Data:

- Certain countries or years in the dataset have missing data, which limits the completeness and accuracy of the analysis.
- The absence of data for some economic indicators can impact the depth of the insights derived.

2. Excluded Factors:

- The relationship between GDP and inflation is influenced by various factors such as trade policies, geopolitical events, and external debt. These were not included in this analysis, which may affect the interpretation of results.

Key Learnings

1. GDP Trends:

- Analyzing GDP trends over time provides valuable insights into periods of economic growth, recovery, and stagnation. It highlights the impact of major global events such as financial crises and pandemics.

2. Global Economic Disparities:

- Comparing GDP across countries showcases the economic disparities between nations, emphasizing the dominance of high-income countries and the challenges faced by low-income nations.

3. Interconnectedness of Indicators:

- Correlation analysis between GDP and inflation reveals relationships that can guide economic policymaking. For instance, higher GDP is often associated with stable inflation in well-developed economies.

Future Improvements 1. Incorporate Additional Datasets: - Expanding the analysis to include other economic indicators such as trade balance, unemployment rates, and external debt can provide a more holistic view of economic performance.

2. Machine Learning for Prediction:

- Utilizing machine learning models can enhance the analysis by predicting GDP growth trends based on historical data and other economic variables.
- This could help policymakers anticipate future challenges and opportunities.

3. Regional and Sectoral Analysis:

- Breaking down GDP and inflation trends by regions or economic sectors could uncover more granular insights and allow for targeted recommendations.

4. Longer Timeframes:

- Analyzing longer historical trends and including projected data beyond 2023 could provide a more comprehensive perspective on economic stability and growth.

The analysis is complete, and the outputs are ready:

1. Cleaned Dataset: A CSV file containing the cleaned data is saved here.

```
[33]: cleaned_data.to_csv('cleaned_weo_data.csv', index=False)
```

2. Visualization: Line Chart The GDP trend of the United States (2000-2023) is visualized as a line chart. The visualization is saved as a PNG image (us_gdp_trend.png).

```
[37]: plt.savefig('us_gdp_trend.png')
```

<Figure size 640x480 with 0 Axes>

References: - International Monetary Fund. (2023). World Economic Outlook Database, October 2023. Retrieved from <https://www.imf.org/en/Publications/WEO/weo-database/2023/October>
- Python Software Foundation. (2023). Pandas Documentation. Retrieved from <https://pandas.pydata.org/>
- Matplotlib Developers. (2023). Matplotlib Documentation. Retrieved from <https://matplotlib.org/>