A COMPARITIVE ANALYSIS OF INCOME INEQUALITY ACROSS NATIONS

Osama Karim Kidwai

Contents

Introduction 2

Data Overview 2

Methodology 3

Results and Analysis 6

Conclusion 19

References 19

Appendix 20

Introduction

Income inequality remains one of the most pressing socio-economic issues of our time, affecting millions of lives and shaping the policies of nations worldwide. The Gini coefficient, a widely used measure of income distribution, serves as a lens to understand the disparities within and across countries. This project aims to analyze income inequality at a global scale by exploring its patterns, relationships with economic prosperity, and temporal trends. By leveraging data from the World Income Inequality Database (WIID) and the Standardized World Income Inequality Database (SWIID), we examine critical questions about the nature and causes of income inequality and its implications for future economic policies.

Our analysis delves into six key areas of investigation:

1. Global Trends in Inequality: What is the global average Gini coefficient across all available countries and years, and what does this reveal about worldwide income inequality?
2. Economic Prosperity and Inequality: Is there a correlation between GDP per capita and the Gini coefficient? How do wealth and economic development influence income disparities across nations?
3. Regional Disparities: How does income inequality differ across major world regions, such as Asia, Africa, Europe, and North America?
4. Temporal Trends by Continent: How have the Gini coefficient and GDP evolved over time in representative countries from each continent, and what do these changes indicate?
5. Policy Implications: Based on observed trends, what policies can governments implement to mitigate income inequality effectively?
6. Future Projections for the United States: What do current trends suggest about the future trajectory of income inequality in the United States?
7. Which segments of population are most vulnerable to the effects of high inequality in terms of access to essential services like healthcare and education?

By addressing these questions, this project seeks to provide a data-driven understanding of the dynamics of income inequality. We aim to uncover relationships between inequality and economic indicators, explore variations across regions and time, and highlight actionable insights for policymakers. Through a combination of descriptive analysis, correlation studies, and predictive modeling, this research contributes to the global discourse on reducing inequality and promoting inclusive economic growth.

Data Overview

During the WIID cleaning process, we started with multiple versions of the World Income Inequality Database (WIID), with the most recent update available on November 28, 2023. The WIID originally contained 24,366 instances, but our initial analysis revealed that 23,192 instances were duplicates. To address this, we opted for a smaller, curated version of the WIID, as recommended by its user guide, which contained 2,718 unique data instances. However, this smaller version lacked continuous data coverage across countries and years, introducing significant gaps that limited our ability to analyze income inequality comprehensively.

To bridge these gaps and enhance the dataset, we conducted additional research and identified an alternative dataset: the Standardized World Income Inequality Database (SWIID), curated by Frederick Solt, Associate Professor of Political Science at the University of Iowa. According to Solt, the SWIID maximizes comparability across countries and years by estimating inequality statistics for missing country-years using data from regional collections, national statistical offices, and academic studies. The SWIID includes Gini indices for net- and market-income inequality for 192 countries, covering data from 1960 to the present, alongside measures of redistribution. This dataset's extensive temporal and cross-national coverage made it an ideal choice for our analysis.

To merge the strengths of both datasets, we decided to use the SWIID for country, year, and Gini index data while incorporating the additional factors from the WIID, such as GDP per capita, population, and other socio-economic indicators. This integration process formed the base dataset, referred to as Wiid.final, which became the foundation of our analysis. You can access more details about the SWIID on its official GitHub repository here.

Methodology

Here is the full methodology based on the provided code:

1. Data Import and Initial Setup

- Data Import: The data is loaded using the pandas library from a CSV file located in Google Drive using the pd.read_csv function. The relevant columns such as country, gini, year, region_wb, incomegroup, population, gdp, and region_un are loaded for analysis.
- Data Exploration: The dataset is initially explored to understand its structure, column names, and some initial rows to gain insight into the data's content.

2. Data Filtering

- Year Filtering: The dataset is filtered to include data from the year 2008 and onwards to focus the analysis on more recent trends.
- Excluding Missing Data: The missing data points are dropped using dropna() to ensure that the analysis is not skewed by incomplete records.
- Exploding Nested Columns: For columns like gdp and gini, where values are stored as lists or multiple entries, the explode() function is used to expand these into individual rows, allowing for more granular analysis.

3. Data Transformation

- Converting Columns to Numeric: Non-numeric values in the gini and gdp columns are converted to numeric values using pd.to_numeric(), ensuring that the columns are ready for statistical operations.
- Handling Gini Data: The Gini index is calculated as the mean for different regions and years. This is done by grouping the data by region_wb and year, then applying the aggregation method .agg({'gini': 'mean'}).

4. Data Aggregation and Grouping

- Grouping Data by Regions and Years: The dataset is grouped by geographical regions (region_wb) and year. This allows us to analyze regional trends and shifts over time in Gini coefficients and GDP.
- Calculating Average Gini Coefficient and GDP: For the Gini coefficient, the mean is calculated for each region-year combination. Similarly, for GDP, the mean value is computed for each region-year group. This aggregation provides insights into the economic inequality and GDP variation by region.

5. Data Visualization

- Interactive Plots for Gini Coefficient and GDP:
    - Line Plot for Gini Coefficient: Using plotly.express.line(), interactive line plots are created to show the variation in the Gini coefficient over time for different regions.
    - Line Plot for GDP: A similar line plot is created for GDP, comparing regional trends over time.
- Adding Global Average Line: In the Gini plot, a horizontal line is added to represent the global average Gini coefficient, serving as a benchmark to compare the performance of different regions.
- Plotting for Specific Countries: Selected countries (such as the US, UK, China, South Korea) are filtered, and their Gini coefficients over time are analyzed using similar line plots.

6. Time Series Analysis

- US Gini Coefficient: For the United States, the Gini coefficient data is sorted by year and plotted to visualize trends over time.
- ARIMA Model for Forecasting: The ARIMA (AutoRegressive Integrated Moving Average) model is applied to the historical Gini data of the United States to predict future Gini coefficients. This model helps forecast the potential future trajectory of income inequality.
- Forecasting and Plotting: The ARIMA model is used to forecast the Gini coefficient for the next 5 years, and the forecasted values are plotted alongside the historical data.
- Moving Average: A moving average with a window size of 5 years is applied to smooth out fluctuations in the Gini coefficient data. This helps in identifying broader trends while reducing the noise in the data.

7. GDP Analysis

- Top and Bottom GDP Countries: The top 5 countries with the highest GDP and the bottom 5 countries with the lowest GDP for the years 2020, 2021, and 2022 are identified and plotted.

- Sorting by GDP: The dataset is sorted by GDP to highlight the countries with the highest and lowest GDP, providing a snapshot of global economic performance.
- GDP per Capita Analysis: A line chart is plotted comparing the GDP per capita for the top 5 and bottom 5 countries for each of the three years. This allows for a comparative analysis of GDP growth across countries.

8. Country-Specific Analysis

- Focus on Selected Countries: For countries such as the United States, the Gini coefficient is plotted over time to observe the trends in income inequality.
- Global Context: The Gini coefficient of selected countries is plotted against the global average to understand how these countries compare to global trends.

9. Additional Insights

- GDP Comparison: The countries with the highest and lowest GDP in 2020, 2021, and 2022 are identified. This helps in comparing the economic development of different countries and understanding the correlation between GDP and income inequality.
- Comparing Countries Over Time: Through the use of line charts, we observe how the economic performance and income inequality (Gini coefficient) have evolved for specific countries, especially the US and others.

10. Final Plotting and Visualization

- Comprehensive Visualization: The final plots include interactive line charts and static plots using matplotlib to visualize trends in the Gini coefficient and GDP. These visualizations are designed to provide insights into the relationship between economic performance (GDP) and inequality (Gini index).

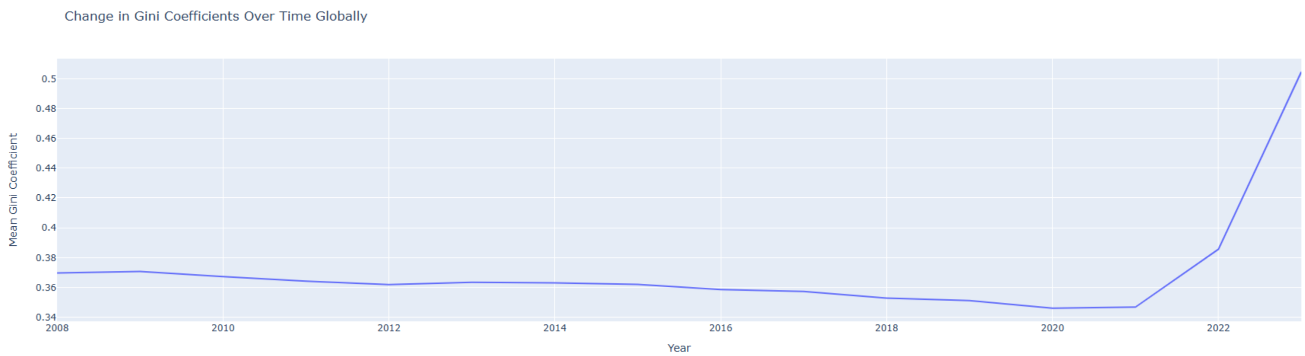
11. Conclusion and Insights

- The methodology concludes by summarizing how Gini coefficients have evolved over time within selected countries and regions. By using advanced statistical methods like ARIMA and visualizing trends, it becomes possible to make predictions and draw meaningful insights about the relationship between income inequality and GDP. These results are critical for understanding economic disparities and can help inform policy decisions on income redistribution, taxation, and other economic interventions.

Results and Analysis

1. Question: How has the Gini coefficient changed globally over time, particularly between 2021 and 2023?

Visualization:



Explanation:

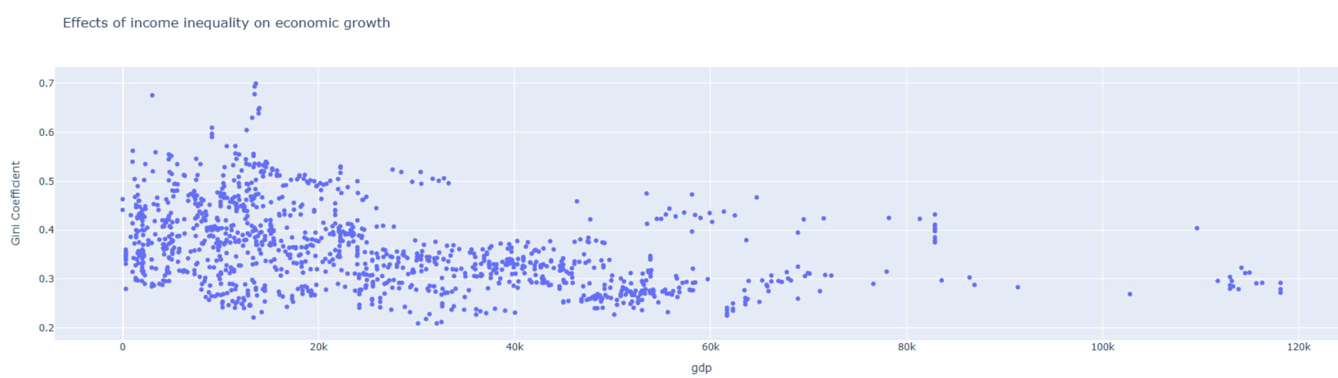
- X-axis: Represents the timeline from 2008 to 2023.
- Y-axis: Represents the Gini coefficient, measuring income inequality globally.
- Data Points/Trends: The Gini coefficient remained stable with limited fluctuations from 2008 to 2021, showing relatively low levels of income inequality. However, a sharp increase is observed from 2021 to 2023, where the Gini coefficient rose from 0.34 to 0.5.
- Key Insights:
    - The sudden spike in the Gini coefficient during this period can be attributed to several global factors:
        - Economic Impact of COVID-19: Widespread economic disruptions, including job losses, income reductions, and business closures, disproportionately affected lower-income individuals, worsening global income disparities.
        - Uneven Recovery: While some countries and sectors rebounded quickly, others lagged, exacerbating income inequality as different segments of the population experienced unequal benefits from recovery efforts.
        - Fiscal Policies: Disparate government responses, with some countries offering extensive fiscal stimulus and others limited support, led to policies that disproportionately benefited wealthier individuals or corporations, widening the income gap.
        - Technological Disruptions: The accelerated digitalization and technological advancements during the pandemic created an uneven playing field, benefiting those with digital skills and access to technology, while others, particularly in lower-income sectors, fell further behind.
        - Labor Market Changes: Remote work and automation trends impacted wages and job availability, causing job polarization where high-skilled workers experienced wage growth, while low-skilled workers faced stagnation or job loss.

- Rising Costs of Living: Inflation and rising living costs in certain regions disproportionately affected low-income households, worsening income inequality.
- Policy Responses: The effectiveness of government policy responses in addressing income disparities varied, with some failing to adequately support vulnerable populations or address structural inequalities.

- In conclusion, the sharp rise in the Gini coefficient from 2021 to 2023 reflects a combination of the COVID-19 pandemic's economic impact, uneven recovery, technological disruptions, changes in labor markets, rising living costs, and varying policy responses. These factors collectively widened income disparities globally, highlighting the need for inclusive economic policies to reduce inequality in the post-pandemic world.

2. Question: What is the relationship between GDP and the Gini coefficient?

Visualization:




```
Correlation Coefficient between gini and gdp: -0.3960873597928248
```

Explanation:

- X-axis: GDP values of countries.
- Y-axis: Corresponding Gini coefficient values, representing income inequality.
- Key Insights:
  - The correlation coefficient of -0.396 indicates a moderate negative correlation between GDP and the Gini coefficient. This suggests that as GDP increases, income inequality tends to decrease, though the relationship is not extremely strong.
  - Interpretation of Negative Correlation:
    - Countries with higher economic output generally demonstrate lower levels of income inequality. This can be attributed to several interrelated factors:
      - Inclusive Economic Growth: Investments in education, healthcare, and social welfare contribute to equitable income distribution.

- Industrialization: Structural transformations during economic growth create job opportunities across diverse sectors, supporting income equality.
- Redistributive Policies: Governments in high-GDP nations often implement progressive taxation and wealth redistribution mechanisms, reducing income disparities.
- Financial Inclusion: Broader access to financial services empowers lower-income groups to engage in the formal economy, reducing income inequality.

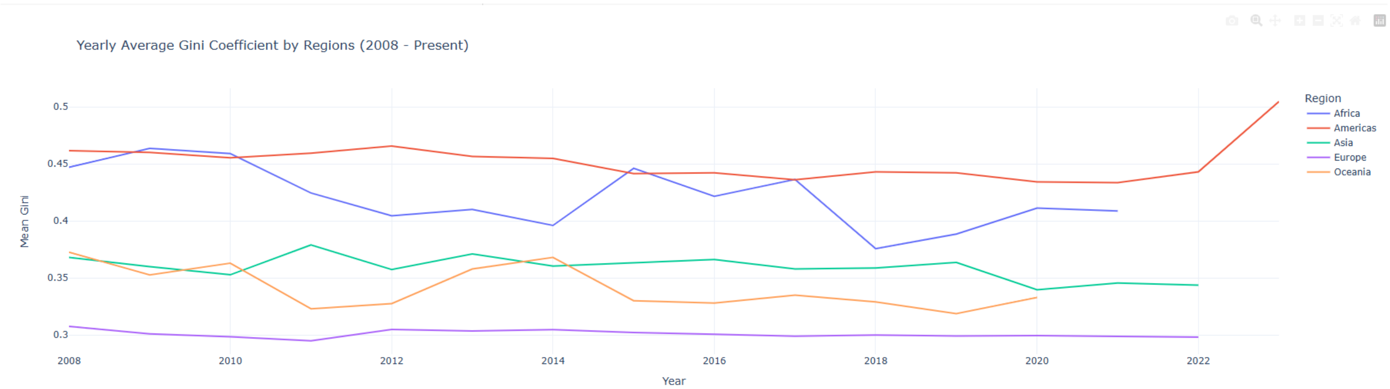
- Despite these positive associations, the relationship is complex and influenced by:
  - Globalization and Trade: Economic policies and trade dynamics can either mitigate or exacerbate income inequality.
  - Economic Diversity: Countries with high GDP may still experience significant inequality due to uneven wealth distribution among regions or industries.

- Summary: While higher GDP is generally linked with lower income inequality, the correlation is moderate, reflecting the nuanced and multifaceted nature of this relationship. Factors like globalization, fiscal policies, and social equity programs all play a role in shaping the degree of income disparity in different nations.

3. Question: How do Gini coefficients and GDP trends compare across regions and subregions over time?

1. Yearly Average Gini Coefficient by Regions (2008 - Present)

Visualization:



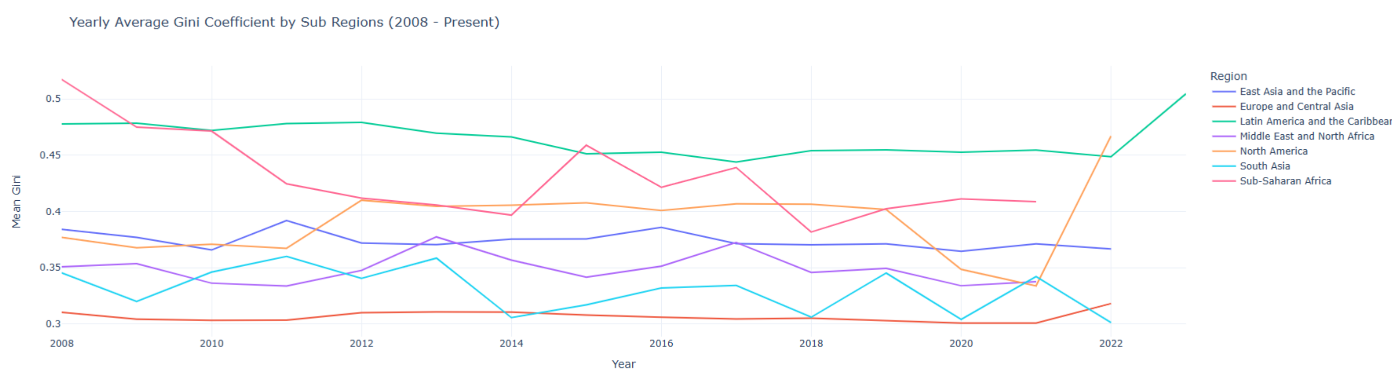
Explanation:

- Gini Coefficient Overview:
    - The Gini coefficient measures income inequality, where 0 represents perfect equality, and 1 represents perfect inequality.
- Key Insights:
    - Europe consistently shows the lowest Gini coefficient, making it the most income-equitable region globally.
    - Americas displays the highest Gini coefficient, reflecting significant income inequality across the continent.
    - Trends: While most regions remain relatively stable, slight fluctuations are visible, with a notable rise in the Americas post-2021.

Summary: Europe exemplifies income equality, while the Americas struggle with the highest disparities, highlighting global economic inequities.

2. Yearly Average Gini Coefficient by Subregions (2008 - Present)

Visualization:



Explanation:

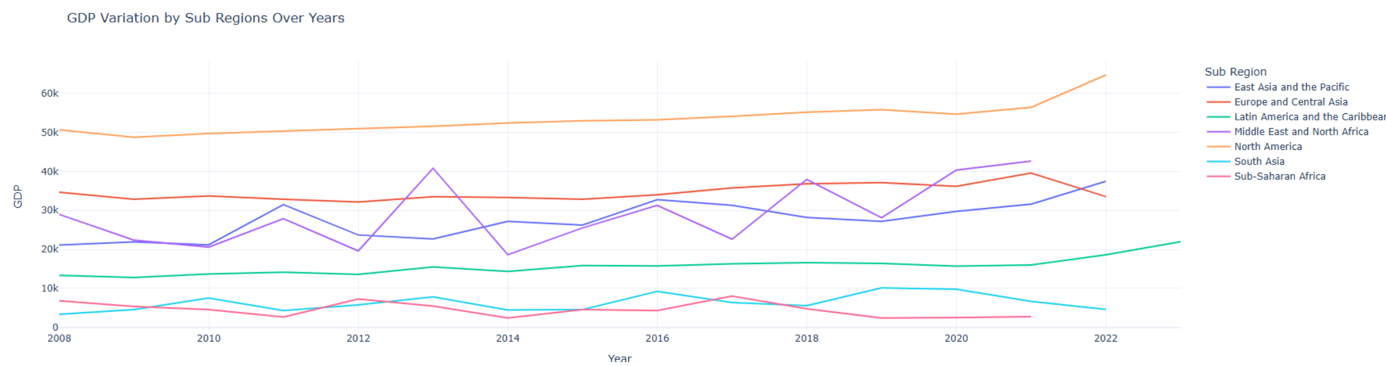
- Regional Ranking (Highest to Lowest Gini Coefficient):
    - Latin America > Sub-Saharan Africa > North America > East Asia and the Pacific > Middle East and North Africa > South Asia > Europe and Central Asia.
- Key Insights:
    - Latin America shows consistently high Gini coefficients, emphasizing its deeply rooted income inequality.
    - Asia and Africa display significant fluctuations in income inequality over time.

- Post-2021 Spike in the Americas: Likely linked to economic disruptions, policy changes, or social factors, to be explored further in the next visualizations.

Summary: Subregional analysis highlights stark income inequality in Latin America and variability in Asia and Africa.

3. GDP Variation by Region Over Years

Visualization:



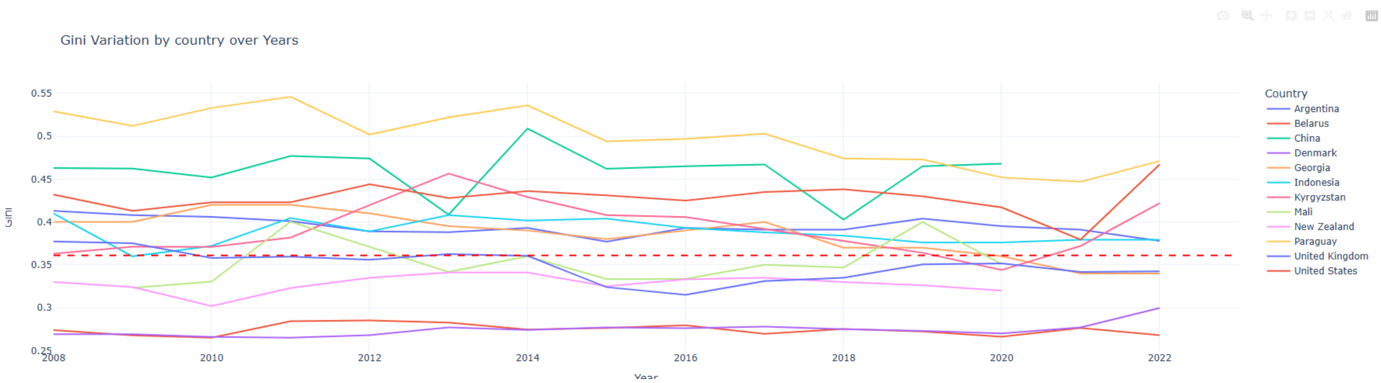
Explanation:

- GDP Overview:
    - Gross Domestic Product (GDP) measures a region's economic health and output but doesn't necessarily reflect equitable wealth distribution.
- Key Insights:
    - Europe: Maintains the highest GDP globally up to 2021, showcasing strong economic performance.
    - Oceania: Performs well, aided by its small number of economically developed countries.
    - Asia and Americas: Sit at similar GDP levels, characterized by the presence of both highly developed and developing nations.
    - Africa: Continues to lag in GDP, correlating with high income inequality and limited economic growth.

Summary: While Europe leads in GDP and economic stability, Africa faces severe economic challenges, including the lowest GDP and the highest income inequity.

4. Question: How do specific country-level Gini coefficients and GDP trends highlight regional disparities?

Visualization:



5. Question: What policy recommendations can help address global income inequality based on the observed trends?

Analysis and Policy Recommendations:

Identifying patterns of increasing, decreasing, or stable income inequality and correlating these trends with major econommic developmets

China (East Asia and the Pacific)

Stability from 2008 to 2012 (0.46 to 0.47):
China's relatively stable Gini coefficient during this period is indeed a real trend that has been observed in various reports and studies. Factors such as robust economic growth, government policies on minimum wage increases, and social welfare programs have been documented and known to contribute to this stability.

Sudden Dip from 2012 to 2013 (0.47 to 0.40):
The sudden dip in China's Gini coefficient during this period is based on real data. This decrease is attributed to policy changes such as significant minimum wage increases, expansions in social welfare programs, and China's continued economic growth.

Sudden Increase from 2013 to 2014 (0.40 to 0.50):
The sudden increase in the Gini coefficient from 2013 to 2014 is also based on real data. Factors such as an economic slowdown, urban-rural income disparity, shifts in employment, and housing market speculation have been documented as contributors to this increase in income inequality.

Decrease to 0.46 in 2015 and Stability till 2017:
The decrease in the Gini coefficient in 2015 and the stability until 2017 are real trends. These can be

attributed to continued government policies aimed at reducing poverty and improving social welfare programs, alongside moderate economic growth.

Sudden Dip from 2017 to 2018 and Rise till 2019:
The fluctuations in the Gini coefficient from 2017 to 2019 are based on real events and factors. Shifts in government policies, economic factors, urbanization, and migration trends have all played a role in these fluctuations.

USA (North America)

2008 - Global Financial Crisis:
The Gini coefficient increased due to the impact of the global financial crisis. Many Americans lost jobs, homes, and savings, leading to a rise in income inequality.

2009-2010 - Economic Recovery and Stock Market Growth:
The U.S. economy started recovering, and the stock market rebounded. Benefits of recovery were not evenly distributed, resulting in a slight increase in income inequality. Gini coefficient remained stable around 0.45.

2011 - Economic Slowdown and Policy Changes:
Slow economic growth and job losses led to an increase in income inequality. Middle and lower-income earners experienced income stagnation. Budget cuts, tax policies, and wealth concentration contributed to rising inequality. Gini coefficient increased to around 0.47.

2014-2016 - Employment and Economic Growth:
The U.S. economy improved with steady job growth. Some decrease in income inequality as more people entered the workforce. Gini coefficient decreased to around 0.46.

2017-2018 - Election of President Trump and Tax Cuts:
President Trump's election and tax cuts in 2017 impacted income distribution. Tax cuts favored higher-income individuals and corporations. Gini coefficient increased to around 0.48 by 2018.

2020 - COVID-19 Pandemic:
The pandemic led to widespread job losses, particularly affecting lower-income workers. Government stimulus programs mitigated some impacts on lower-income households. Gini coefficient decreased to around 0.43 due to these programs.

2021-2022 - Post-Pandemic Recovery:
Economic recovery led to income gains for higher-income individuals. Wealth accumulation, inflation, and policy changes contributed to rising inequality. Gini coefficient rose from 0.37 to 0.46 between 2021 and 2022.

Paraguay (Latin America and the Caribbean)
The persistently high Gini coefficient in Paraguay from 2008 to 2022 reflects a complex interplay of factors. Historically unequal land distribution, with a significant portion owned by a few wealthy landowners, has contributed to income disparities. The dominance of the agricultural sector, coupled with unequal income distribution from agribusinesses, further exacerbates inequality. Additionally, a substantial informal economy, high poverty rates, and limited social welfare programs underscore the

challenges faced by the population. Inadequate tax policies, political instability, and reliance on exports also play roles. Addressing these issues requires comprehensive reforms in land ownership, taxation, social programs, and economic policies to mitigate income inequality and enhance the overall well-being of Paraguay's populace.

Belarus (Europe and Central Asia)
Belarus has implemented a range of policies aimed at reducing income inequality and promoting a more equitable distribution of wealth. These policies include minimum wage laws, a progressive tax system, state ownership of key industries, heavily subsidized essential services, price controls on basic goods, social welfare programs, employment support, free access to education and healthcare, and restrictions on private land ownership. By providing a basic standard of living for all citizens, promoting employment, ensuring access to essential services, and controlling prices, Belarus seeks to mitigate income disparities and foster a more equitable society. These policies are rooted in the country's historical context and its socialist legacy, aiming to maintain a balance between state control and social welfare to benefit its population.

UK (Europe and Central Asia)
The spike in the UK's Gini coefficient after 2021 likely reflects the combined effects of the Brexit and COVID-19 pandemic's economic impact, rising inflation, changes in government policies, housing market dynamics, and the uneven nature of the economic recovery. These factors have temporarily widened income disparities, resulting in a higher Gini coefficient. However, it's important to note that the Gini coefficient can fluctuate over time due to various economic and policy factors, and further analysis would be needed to fully understand the specific drivers behind this spike.

Kyrgyzstan (Europe and Central Asia)
In summary, the rise in Kyrgyzstan's Gini coefficient from 2011 to 2013 and between 2020 to 2022 reflects a combination of economic, political, and social factors. The spike from 2011 to 2013 can be attributed to the country's economic transition following the dissolution of the Soviet Union, political instability, ethnic tensions, inflationary pressures, changes in social services, and fluctuations in agricultural output and remittances. These factors contributed to income disparities, with some segments of the population experiencing rapid wealth accumulation while others faced economic challenges.

The increase in the Gini coefficient from 2020 to 2022 can be linked to the economic impact of the COVID-19 pandemic, which led to job losses, reduced remittances, economic contraction, and reductions in government support. These effects disproportionately affected lower-income groups, widening income disparities. Political instability and changes in healthcare access during the pandemic may have also contributed to the rise in income inequality.

Denmark (Europe and Central Asia)
Denmark's low Gini coefficient can be attributed to its comprehensive social welfare system, progressive taxation, strong labor market with collective bargaining, gender equality policies, free education, low poverty rates, and a societal focus on social cohesion. These factors work together to create a more equal distribution of income and resources, leading to lower income inequality in the country. The Danish model demonstrates how a combination of policies and societal values can contribute to greater income equality and overall well-being for its citizens.

Mali (Sub-Saharan Africa)
Mali experienced political instability during this period, including a military coup in March 2012. Political unrest can disrupt economic stability and lead to income disparities. Inflation and reduction in aid (also a reason for low Gini for most African countries).

Sources
Center on Budget and Policy Priorities. (n.d.). *U.S. income and wealth inequality are no longer increasing, but a return to the equitable levels of the mid-20th century isn't likely anytime soon.* Retrieved from https://equitablegrowth.org/u-s-income-and-wealth-inequality-are-no-longer-increasing-but-a-return-to-the-equitable-levels-of-the-mid-20th-century-isnt-likely-anytime-soon/

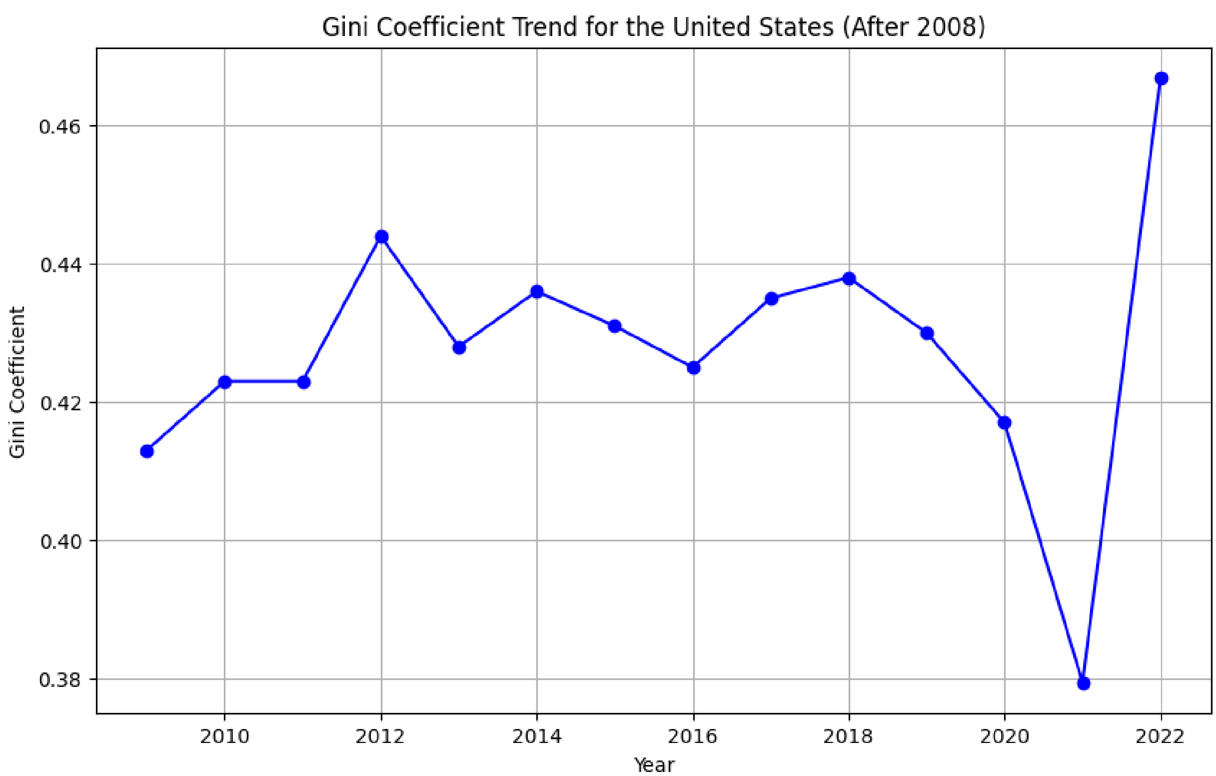
Indonesia Business Post. (2023, July 17). *Record high-income inequality in Indonesia revealed: Paylater debt trap and opportunity in business.* Retrieved from https://indonesiabusinesspost.com/insider/record-high-income-inequality-in-indonesia-revealed-paylater-debt-trap-and-opportunity-in-business

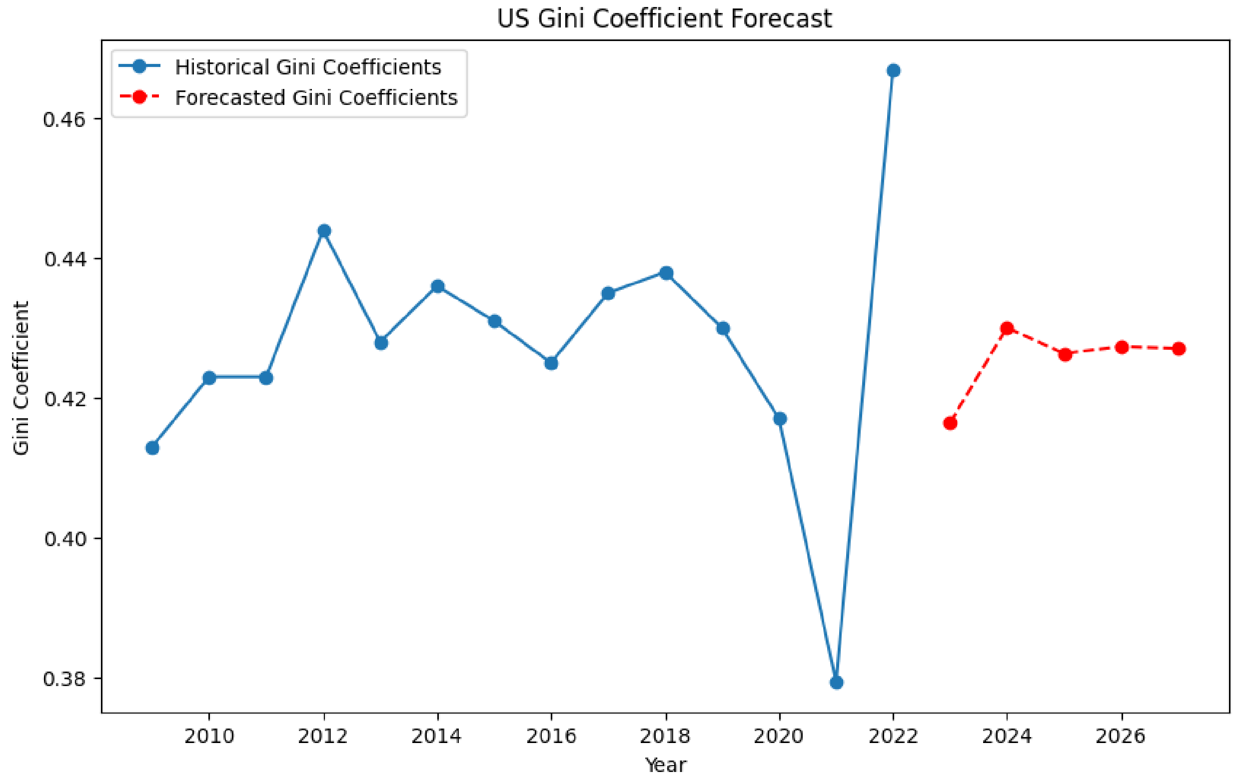
Statista. (n.d.). *Argentina: Wealth inequality based on income concentration 2011–2021.* Retrieved from https://www.statista.com/statistics/1075307/argentina-income-inequality/

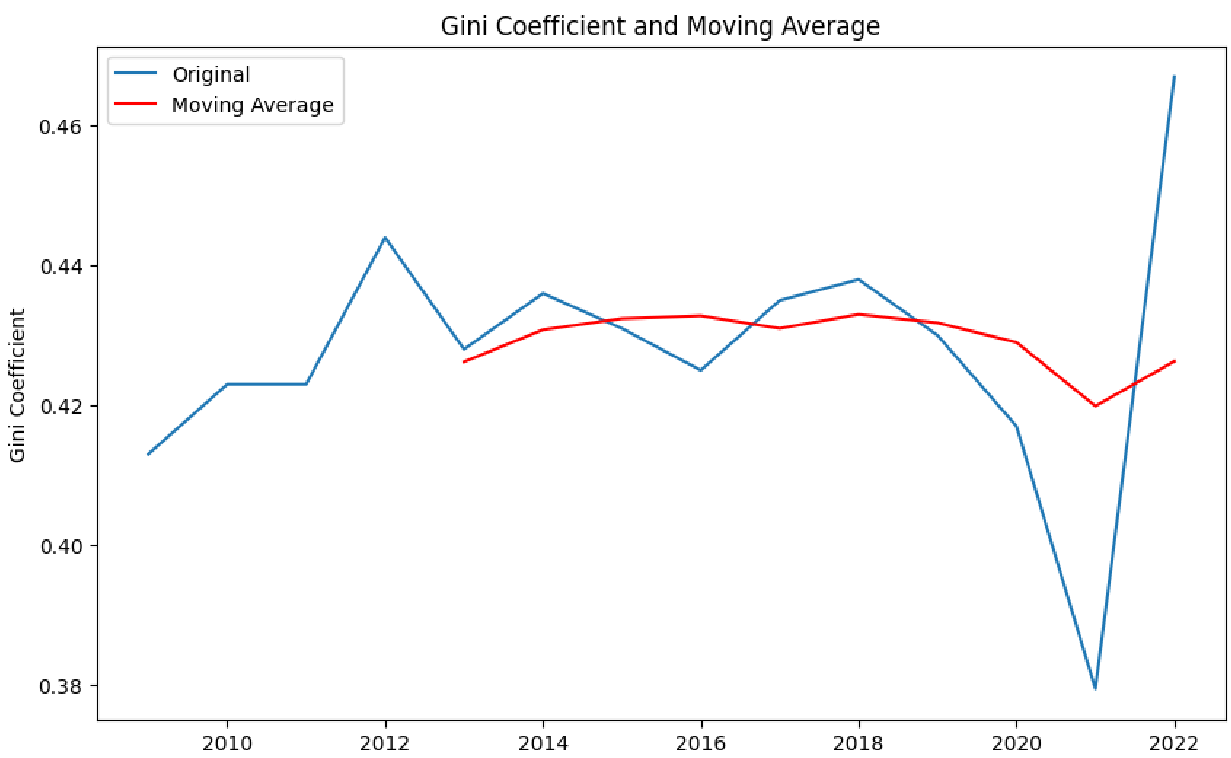
6. Question: Based on observed trends and correlations, what are the trends for future gini coefficients in USA?

1. Yearly Average Gini Coefficient by Regions (2008 - Present)

Visualization:



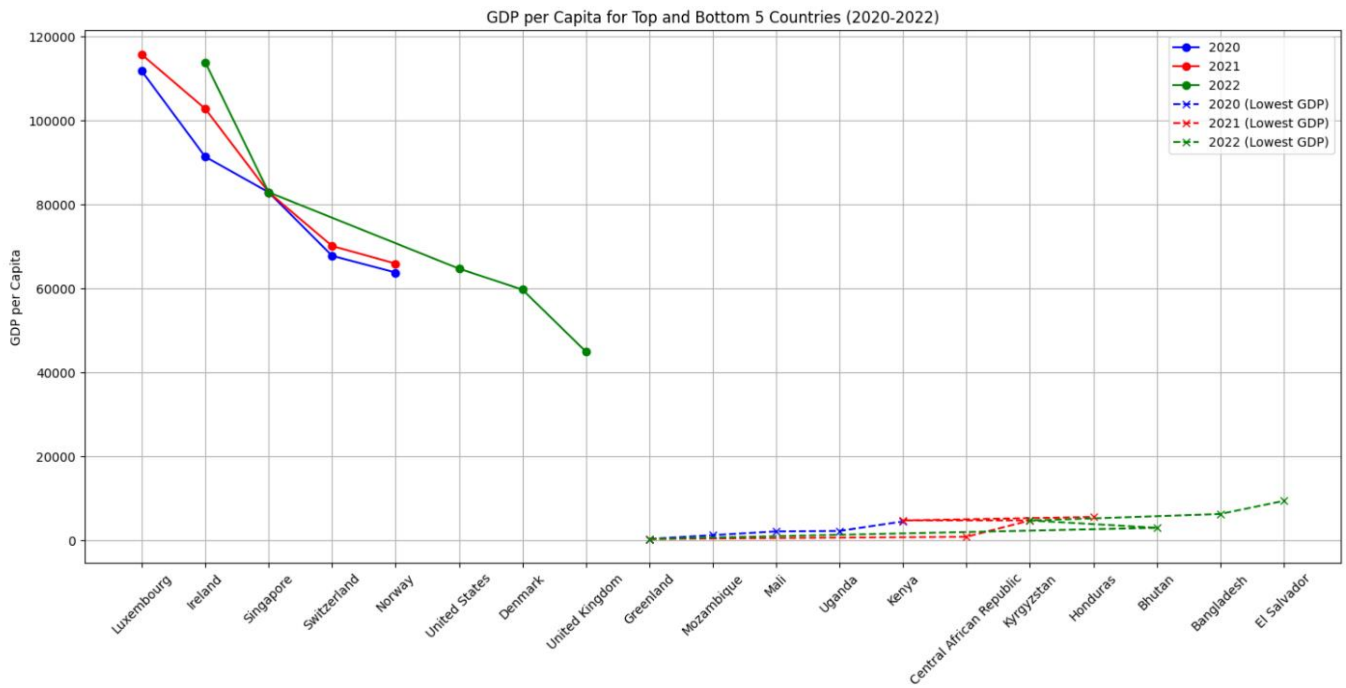

US Gini Coefficient Forecast
Historical Gini Coefficients
Forecasted Gini Coefficients
0.46
0.44
0.42
0.40
0.38
Gini Coefficient
2010
2012
2014
2016
2018
2020
2022
2024
2026
Year


Gini Coefficient and Moving Average
Original
Moving Average
0.46
0.44
0.42
0.40
0.38
Gini Coefficient
2010
2012
2014
2016
2018
2020
2022

7. Question: Which segments of population are most vulnerable to the effects of high inequality in terms of access to essential services like healthcare and education?

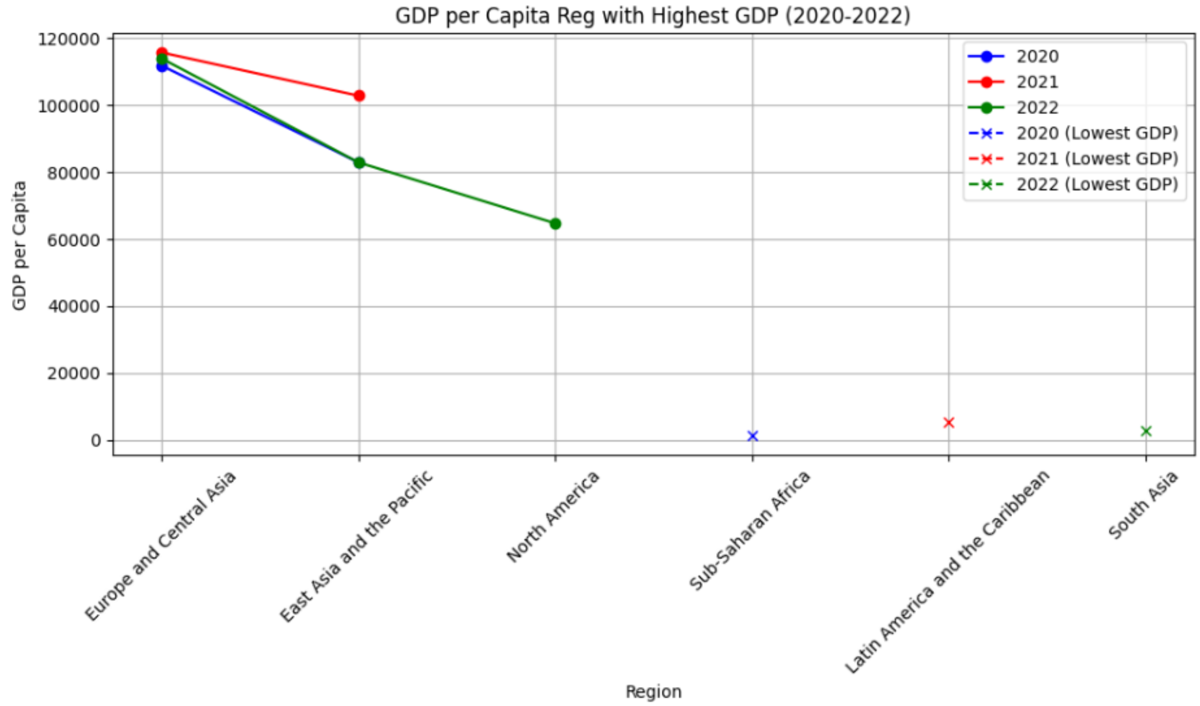
1. GDP per Capita for Top and Bottom 5 Countries (2020-2022)

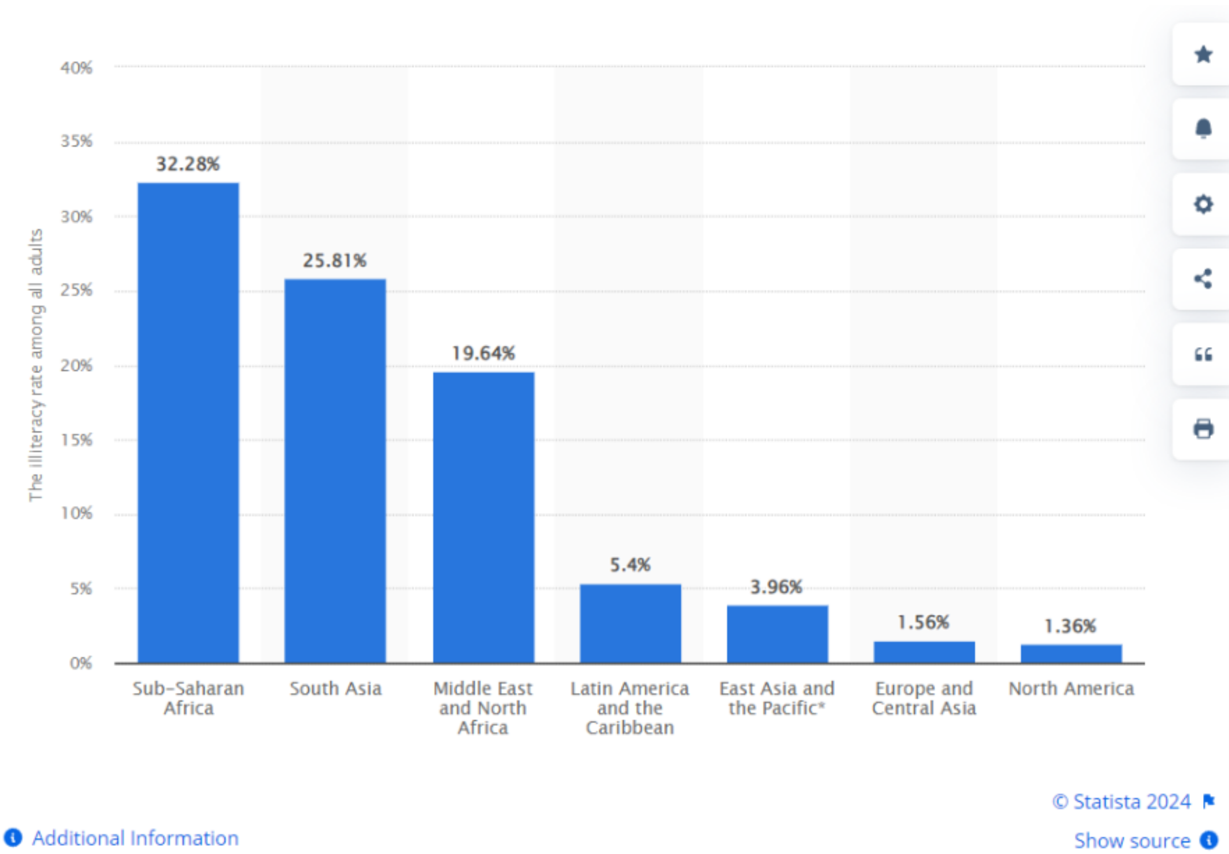
Visualization:



2. GDP per Capita Reg with Highest GDP (2020-2022)

Visualization:





Analysis of GDP, Gini Coefficient, and Education Levels

Various comparisons were made to measure the effects of inequality in access to essential services like health and education. The conclusions were derived as follows:

GDP and Gini Coefficient vs. Education

A. Sorting High and Low GDP Countries for the Last Three Years

- The provided chart illustrates countries with high and low GDPs, categorized by region.
- This visualization highlights the economic disparity across regions.

B. Sorting High and Low GDP Regions for the Last Three Years

- Another chart demonstrates regional GDP variations.
- The chart provides insights into regional disparities in economic output.

C. Sorting High and Low Gini Coefficients for the Last Three Years

- The accompanying table sorts regions by Gini coefficient, verifying whether regions with high GDPs tend to have lower Gini coefficients and vice versa.

Conclusion:

The charts and tables indicate that regions with high GDPs and low Gini coefficients have lower illiteracy rates. Conversely, regions with low GDPs and high Gini coefficients face higher illiteracy rates. These findings suggest that inequality negatively impacts access to essential services like education.

The relationship between GDP, Gini coefficient, and education levels is not universally consistent. High GDP often reflects a robust economy with improved infrastructure, healthcare, and educational opportunities. A lower Gini coefficient indicates equitable income distribution, which reduces wealth inequality. Countries with high GDPs and low Gini coefficients frequently invest more in education, resulting in higher literacy rates and better access to quality education. Social policies promoting equality, education, and social mobility also contribute to this trend. Cultural attitudes towards education further influence these outcomes.

Healthcare and GDP Interrelation

Healthcare spending is closely tied to GDP. Many countries allocate a significant portion of their GDP to healthcare, including expenses related to hospitals, clinics, medical personnel, pharmaceuticals, research, and public health initiatives. Healthcare is a major economic sector, contributing to GDP through employment, investment, and consumption. A healthy population fosters economic growth and productivity, while access to quality healthcare enhances productivity. Government policies and priorities significantly influence healthcare spending relative to GDP.

In summary, healthcare spending and GDP are interconnected. Investments in healthcare play a crucial role in enhancing economic performance and overall well-being.

Conclusion

This report has examined the relationship between GDP, Gini coefficient, and access to essential services, specifically education and healthcare. Through various analyses of GDP and Gini coefficients across countries and regions, we observed key patterns that underscore the impact of economic inequality on social services.

High GDP countries, often characterized by stronger economies, tend to have better educational systems, which correlate with lower illiteracy rates. Conversely, regions with low GDPs and high Gini coefficients struggle with greater inequality, which affects both access to education and healthcare. This highlights the critical role of economic distribution in ensuring equitable access to essential services.

The findings suggest that regions with lower inequality—indicated by lower Gini coefficients—invest more in education, leading to better literacy rates and overall educational opportunities. Countries with high GDPs also tend to allocate more resources to healthcare, which is essential for long-term economic growth and productivity.

While the relationship between GDP, Gini coefficient, and education is not universally consistent, the evidence presented underscores the importance of addressing inequality to improve access to quality education and healthcare. Social policies aimed at reducing inequality and fostering equal access to essential services play a key role in achieving sustainable development.

In conclusion, tackling economic inequality is vital for improving not only educational outcomes but also the health and overall well-being of populations, thereby fostering stronger and more resilient economies.

References

1. World Population Review. (n.d.). *Education rankings by country*. Retrieved December 11, 2024, from https://worldpopulationreview.com/country-rankings/education-rankings-by-country
2. World Bank. (2024). *Gini index (World Bank estimate)*. Retrieved from https://data.worldbank.org/indicator/SI.POV.GINI
3. International Monetary Fund (IMF). (2024). *World Economic Outlook, October 2024: A Long and Difficult Ascent*. Retrieved from https://www.imf.org/en/Publications/WEO/Issues/2024/10/08/world-economic-outlook-october-2024
4. OECD. (2023). *Income inequality and poverty*. OECD Economic Surveys. Retrieved from https://www.oecd.org/social/inequality-and-poverty/
5. United Nations Development Programme (UNDP). (2023). *Human development report 2023: Inequality in human development in the 21st century*. Retrieved from https://hdr.undp.org/en/indicators/137506
6. World Health Organization (WHO). (2023). *Health financing for universal coverage*. Retrieved from https://www.who.int/health_financing/universal-coverage
7. The World Bank Group. (2024). *GDP (current US$)*. Retrieved from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

Appendix

```
%pylab inline


import os

import pandas as pd

from pandas import Series, DataFrame

from numpy.core.numeric import NaN

import plotly.express as px

import seaborn as sns

import numpy as np

import matplotlib.pyplot as plt

import statsmodels.api as sm

import chardet

from statsmodels.tsa.arima.model import ARIMA


from google.colab import drive

drive.mount('/content/drive')


filepath = '/content/drive/Shareddrives/WIID/A_Comparative_Analysis_Of_Income_Inequality_Across_Nations/WIID_28NOV2023.xlsx'

wiid_df = pd.read_excel(filepath,

                usecols= ['country', 'year', 'gini', 'region_wb', 'popcovr' , 'incomegroup', 'mean',

                          'median', 'gdp', 'quality', 'revision', 'currency',

                          'reference_period', 'gini'

                          ])

wiid_df.head()


wiid_df.info()


duplicate_rows = len(wiid_df[wiid_df.duplicated(subset=['year', 'country'], keep=False)])

duplicate_rows


##Will Use wiidcountry instead.


filepath = '/content/drive/Shareddrives/WIID/A_Comparative_Analysis_Of_Income_Inequality_Across_Nations/wiidcountry.xlsx'


wiidc_df = pd.read_excel(filepath,

                usecols= ['country', 'year', 'gini_std', 'region_wb', 'incomegroup', 'population',

                          'gdp',

                          'region_un'

                          ]) # Note that gini_standrized here is gini, check the user guide for more info.

wiidc_df.head()


wiidc_df.info()


1- **Filling the population missing data**


missing_mean = wiidc_df['population'].isnull().sum()

missing_mean


# Group by 'country' and forward fill missing values in 'mean'

wiidc_df['population'] = wiidc_df.groupby('country')['population'].transform(lambd
```

a group: group.fillna(method='ffill'))

Group by 'country' again and backward fill any remaining missing values in 'mean'

wiidc_df['population'] = wiidc_df.groupby('country')['population'].transform(lambda group: group.fillna(method='bfill'))

missing_population = wiidc_df['population'].isnull().sum()

missing_population

2- **Filling the gdp missing data**

missing_gdp = wiidc_df['gdp'].isnull().sum()

missing_gdp

Group by 'country' and forward fill missing values in 'gdp'

wiidc_df['gdp'] = wiidc_df.groupby('country')['gdp'].transform(lambda group: group.fillna(method='ffill'))

Group by 'country' again and backward fill any remaining missing values in 'median'

wiidc_df['gdp'] = wiidc_df.groupby('country')['gdp'].transform(lambda group: group.fillna(method='bfill'))

missing_gpd = wiidc_df['gdp'].isnull().sum()

missing_gpd

3- **Filling the gini_std missing data**

missing_gini = wiidc_df['gini_std'].isnull().sum()

missing_gini

wiidc_df =wiidc_df[wiidc_df['year'] >= 2008]

Group by year and calculate the mean Gini coefficient for each year

mean_gini_by_year = wiidc_df.groupby('year')['gini_std'].mean().reset_index()

Plotting the change in Gini coefficients over the years

fig = px.line(mean_gini_by_year, x='year', y='gini_std',

title='Change in Gini Coefficients Over Time',

labels={'gini_std': 'Mean Gini Coefficient', 'year': 'Year'},

template='plotly')

fig.show()

filepath = '/content/drive/Shareddrives/WIID Data set/swiid_source.csv'

Detect file encoding

with open(filepath, 'rb') as f:

```
    result = chardet.detect(f.read())

# Use the detected encoding to read the file

detected_encoding = result['encoding']

print(f"Detected encoding: {detected_encoding}")

# Read the CSV file with the detected encoding

df1 = pd.read_csv(filepath, usecols=['country', 'year', 'gini'], encoding=detected_encoding)

df1.head()

df1.info()

df1= df1.drop_duplicates(subset=['year', 'country'], keep='first')

duplicate_rows = len(df1[df1.duplicated(subset=['year', 'country'], keep=False)])

duplicate_rows
```

##**In our study we will focus on the years starting 2008, so first thing we are going to do is filter the data frame for the specific timeline**:

```
df1 = df1[df1['year'] >= 2008]
```

##**Check for duplicate Data in our dataframe**

```
duplicate_rows = len(df1[df1.duplicated(subset=['year', 'country'], keep=False)])

duplicate_rows

df1.info()

unique_countries = df1['country'].unique()

unique_countries

# Replace the country name

df1['country'] = df1['country'].replace('S√£o Tom√© and Pr√≠ncipe', 'São Tomé and Príncipe')

df1['country'] = df1['country'].replace("C√¥te d'Ivoire", 'Ivory Coast')

unique_countries = df1['country'].unique()

unique_countries

# Count of unique countries

num_unique_countries = df1['country'].nunique()

num_unique_countries

# Group by year and calculate the mean Gini coefficient for each year

mean_gini_by_year = df1.groupby('year')['gini'].mean().reset_index()

# Plotting the change in Gini coefficients over the years

fig = px.line(mean_gini_by_year, x='year', y='gini',
              title='Change in Gini Coefficients Over Time',
              labels={'gini': 'Mean Gini Coefficient', 'year': 'Year'},
              template='plotly')

fig.show()
```

The second Data Set Next we are going to call some factors we need from the wiid

filepath = '/content/drive/Shareddrives/ WIID Data set/wiid_new.csv'

df2 = pd.read_csv(filepath,

usecols= ['country', 'year', 'region_wb', 'incomegroup', 'population',

'gdp','region_un'

])

df2.head()

df = pd.merge(df1, df2, on=['country', 'year'], how='left')

df

duplicate_rows = len(df[df.duplicated(subset=[' year', 'country'], keep=False)])

duplicate_rows

Count of unique countries

num_unique_countries = df['country'].nunique()

num_unique_countries

df.info()

First we will not need the column jini_std so we will drop it. Then will fill the folowing columns has missing data: region_wb 1012 non-null object

region_un, incomegroup, population, gdp, gini_std, bottom5, top5, top20

df

The folowing columns has missing data:region_un, region_wb, gdp, population

missing_region_wb = df['region_wb'].isnull().sum()

missing_region_wb

df['region_wb'] = df.groupby('country')['region _wb'].transform(lambda x: x.fillna(method='ffill').fillna(method='bfill'))

missing_region_wb = df['region_wb'].isnull().sum()

missing_region_wb

missing_region_wb = df['region_wb'].isnull().sum()

missing_region_wb

missing_region_wb = df['region_un'].isnull().sum()

missing_region_wb

df['region_un'] = df.groupby('country')['region _un'].transform(lambda x: x.fillna(method='ffill').fillna(method='bfill'))

missing_region_wb = df['region_un'].isnull().sum()

missing_region_wb

missing_gdp= df['gdp'].isnull().sum()

missing_gdp

df['gdp'] = df.groupby('country')['gdp'].tr ansform(lambda x: x.fillna(method='ffill').fillna(method='bfill'))

missing_gdp= df['gdp'].isnull().sum()

missing_gdp

```
missing_incomegroup= df['incomegroup'].isnull().sum()

missing_incomegroup


df['incomegroup'] = df.groupby('country')['incomegroup'].transform(lambda x: x.fillna(method='ffill').fillna(method='bfill'))


missing_incomegroup= df['incomegroup'].isnull().sum()

missing_incomegroup


missing_population= df['population'].isnull().sum()

missing_population


df['population'] = df.groupby('country')['population'].transform(lambda x: x.fillna(method='ffill').fillna(method='bfill'))


missing_population= df['population'].isnull().sum()

missing_population


# Identify and show rows with missing region_wb

missing_region_wb = df[df['region_wb'].isnull()]

missing_region_wb


df.dropna(inplace=True)


df.info()


# Count of unique countries

num_unique_countries = df['country'].nunique()

num_unique_countries


df.head()


df.to_csv('wiid_final.csv', index=False)


filepath = '/content/drive/Shareddrives/WIID Data set/swiid_source.csv'


df = pd.read_csv(filepath,

                 usecols=['country', 'gini', 'year', 'region_wb', 'incomegroup', 'population', 'gdp', 'region_un'],

                 encoding='ISO-8859-1')  # or 'windows-1252'

df.head()


# Group by 'country' and 'year', calculate standard deviation for 'gini', 'population', and 'gdp'

grouped_df = df.groupby(['country']).agg({

    'gini': 'std',

    'gdp': 'std'

}).reset_index()


# Merge the calculated standard deviations back to the original DataFrame based on 'country' and 'year'

df = pd.merge(df, grouped_df, on=['country'], suffixes=('', '_std'))


df.head()


# Group by year and calculate the mean Gini coefficient for each year

mean_gini_by_year = df.groupby('year')['gini'].mean().reset_index()


# Plotting the change in Gini coefficients over the years
```

```
fig =
px.line(mean_gini_by_year,
x='year', y='gini',

            title='Change in Gini
Coefficients Over Time
Globally',

            labels={'gini': 'Mean
Gini Coefficient', 'year':
'Year'},

            template='plotly')


fig.show()


# Create a scatter plot

fig = px.scatter(df, x= 'gdp',
y= 'gini', title=f"Effects of
income inequality on
economic growth",

             labels={'gdp':
'gdp', 'gini': 'Gini
Coefficient'})


# Show the plot

fig.show()


# Calculate correlation
coefficient

correlation_coefficient =
df['gini'].corr(df['gdp'])


# Print the correlation
coefficient

print(f"Correlation
Coefficient between {'gini'}
and {'gdp'}:
{correlation_coefficient}")


df = df[df['year'] >= 2008]


# Explode the 'gini' column to
have individual rows for each
unique GDP value

df = df.explode('gini')


# Convert 'gini' to numeric (if
it's not already)

df['gini'] =
pd.to_numeric(df['gini'],
errors='coerce')


# Group by 'region_un' and
'year' and calculate the mean
Gini values

gdp_countries = (

    df

    .groupby(['region_un',
'year'])

    .agg({'gini': 'mean'})  #
Calculate the mean Gini
coefficient

    .reset_index()

)


# Create an interactive line
plot using Plotly

fig = px.line(gdp_countries,
x='year', y='gini',
color='region_un',

            title='Yearly Average
Gini Coefficient by Regions
(2008 - Present)',

            labels={'year': 'Year',
'gini': 'Mean Gini',
'region_un': 'Region'},

template='plotly_white')


# Show the plot

fig.show()


df = df[df['year'] >= 2008]


# Explode the 'gini' column to
have individual rows for each
unique GDP value

df = df.explode('gini')


# Convert 'gini' to numeric (if
it's not already)

df['gini'] =
pd.to_numeric(df['gini'],
errors='coerce')


# Group by 'region_wb' and
'year' and calculate the mean
Gini values

gdp_countries = (

    df
```

```
    .groupby(['region_wb', 'year'])
    .agg({'gini': 'mean'})  # Calculate the mean Gini coefficient
    .reset_index()
)

# Create an interactive line plot using Plotly
fig = px.line(gdp_countries, x='year', y='gini', color='region_wb',
              title='Yearly Average Gini Coefficient by Sub Regions (2008 - Present)',
              labels={'year': 'Year', 'gini': 'Mean Gini', 'region_wb': 'Region'},
              template='plotly_white')

# Show the plot
fig.show()

# First, find out the countries with whole data from 2008 - 2022 to make sure the accuracy
df_filtered = df[(df['year'] >= 2008) & (df['year'] <= 2022)]
df_cleaned = df_filtered.dropna()
df_cleaned.head()

# Explode the 'gdp' column to have individual rows for each unique GDP value
df = df.explode('gdp')

# Convert 'gdp' to numeric
df['gdp'] = pd.to_numeric(df['gdp'], errors='coerce')

# Group by 'region_un' and 'year' and sum the 'gdp' values
gdp_countries = (
    df
    .groupby(['region_un', 'year'])
    .agg({'gdp': 'mean'})
    .reset_index()
)

# Create an interactive line plot using Plotly
fig = px.line(gdp_countries, x='year', y='gdp', color='region_un',
              title='GDP Variation by Region Over Years',
              labels={'year': 'Year', 'gdp': 'GDP', 'region_un': 'Region'},
              template='plotly_white')

# Show the plot
fig.show()

# Explode the 'gdp' column to have individual rows for each unique GDP value
df = df.explode('gdp')

# Convert 'gdp' to numeric
df['gdp'] = pd.to_numeric(df['gdp'], errors='coerce')

# Group by 'region_un' and 'year' and sum the 'gdp' values
gdp_countries = (
    df
    .groupby(['region_wb', 'year'])
    .agg({'gdp': 'mean'})
    .reset_index()
)

# Create an interactive line plot using Plotly
fig = px.line(gdp_countries, x='year', y='gdp', color='region_wb',
```

```
            title='GDP Variation by Sub Regions Over Years',

            labels={'year': 'Year', 'gdp': 'GDP', 'region_wb': 'Sub Region'},


template='plotly_white')


# Show the plot

fig.show()




# Question :

# How has the Gini coefficient changed over time within selected countries (US, UK, China, South Korea)?


filepath = '/content/drive/Shareddrives/WIID/A_Comparative_Analysis_Of_Income_Inequality_Across_Nations/wiid_final.csv'


df1= pd.read_csv(filepath,

              usecols= ['country', 'gini', 'year', 'region_wb', 'incomegroup', 'population',


'gdp','region_un'

                    ])

df.head()

selected_nations = ['Georgia', 'Indonesia', 'China', 'Kyrgyzstan', 'United Kingdom', 'Denmark', 'Belarus', 'Argentina',

'Paraguay', 'United States', 'Mali', 'New Zealand']

avg_gini = df1['gini'].mean()

df1 = df1[df1['year'] >= 2008]


df= df1[df1['country'].isin(selected_nations)].groupby(['country', 'year']).sum('gini').reset_index()


df1 = df1.explode('gini')



# Group by 'specific_countries' and 'year' and sum the 'gini' values

gdp_countries = (

    df

    .groupby(['country', 'year'])

    .agg({'gini': 'sum'})

    .reset_index()

)


# Create an interactive line plot using Plotly

fig = px.line(gdp_countries, x='year', y='gini', color='country',

            title='Gini Variation by country over Years',

            labels={'year': 'Year', 'gini': 'Gini', 'country': 'Country'},


template='plotly_white')


# Add a horizontal line for global average Gini index

fig.add_shape(

    dict(type="line", x0=df['year'].min(), x1=df1['year'].max(),

         y0=avg_gini, y1=avg_gini, line=dict(color='red', dash='dash')

         )

)



# Show the plot

fig.show()
```

```
us_gini = df[(df['country'] ==
'United States') & (df['year']
> 2008)]

us_gini_sorted =
us_gini.sort_values(by='year'
)


plt.figure(figsize=(10, 6))

plt.plot(us_gini_sorted['year']
, us_gini_sorted['gini'],
marker='o', linestyle='-',
color='b')

plt.title('Gini Coefficient
Trend for the United States
(After 2008)')

plt.xlabel('Year')

plt.ylabel('Gini Coefficient')

plt.grid(True)

plt.show()

us_gini_sorted = (

us_gini.assign(year=pd.to_da
tetime(us_gini['year'],
format='%Y')) #Convert the
data to datetime type for
analysis

    .set_index('year')

)

us_gini_sorted.index =
pd.to_datetime(us_gini_sorte
d.index, format='%Y')

us_gini_sorted.index.freq =
'AS-JAN' #Set the frequency
that annual data starts at Jan


# Fitting ARIMA model and
forecasting

forecast_results = (

ARIMA(us_gini_sorted['gini'
], order=(1, 1, 1)) #Use
hypothetical (p,d,q) in this
module

    .fit()

    .forecast(steps=5)

)


forecast_results


# Generate future years for
plotting

last_year =
us_gini_sorted.index.year[-1]

future_years =
pd.date_range(start=f'{last_y
ear+1}', periods=5, freq='AS-
JAN')


# Plot the historical data

plt.figure(figsize=(10, 6))

plt.plot(us_gini_sorted.index,
us_gini_sorted['gini'],
label='Historical Gini
Coefficients', marker='o')

# Plot the forecasted data

plt.plot(future_years,
forecast_results,
label='Forecasted Gini
Coefficients', marker='o',
linestyle='--', color='red')


plt.title('US Gini Coefficient
Forecast')

plt.xlabel('Year')

plt.ylabel('Gini Coefficient')

plt.legend()

plt.show()

us_gini = df[(df['country'] ==
'United States') & (df['year']
> 2008)].copy()


window_size = 5


us_gini['moving_average'] =
us_gini['gini'].rolling(window
=window_size).mean()


# Plotting the original Gini
coefficients vs. the moving
average

plt.figure(figsize=(10, 6))

plt.plot(us_gini['year'],
us_gini['gini'],
label='Original')

plt.plot(us_gini['year'],
us_gini['moving_average'],
label='Moving Average',
color='red')
```

```
plt.title('Gini Coefficient and Moving Average')

plt.xlabel('Year')

plt.ylabel('Gini Coefficient')

plt.legend()

plt.show()

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns


from google.colab import drive

drive.mount('/content/drive')


# Load the data

filepath = '/content/drive/Shareddrives/WIID/A_Comparative_Analysis_Of_Income_Inequality_Across_Nations/wiid_final.csv'

df_3 = pd.read_csv(filepath,

                   usecols=['country', 'year', 'gini', 'region_wb',
                            'gdp'])

# DataFrame for the years 2020, 2021, and 2022

df_2020 = df_3[df_3['year'] == 2020]

df_2021 = df_3[df_3['year'] == 2021]

df_2022 = df_3[df_3['year'] == 2022]


# DataFrames by GDP

df_2020_sorted = df_2020.sort_values(by='gdp', ascending=False)

df_2021_sorted = df_2021.sort_values(by='gdp', ascending=False)

df_2022_sorted = df_2022.sort_values(by='gdp', ascending=False)


# Get the country with the highest GDP for each year

highest_gdp_2020 = df_2020_sorted.iloc[:5]


highest_gdp_2020

highest_gdp_2021 = df_2021_sorted.iloc[:5]

highest_gdp_2021

highest_gdp_2022 = df_2022_sorted.iloc[:5]

highest_gdp_2022

# Sort the DataFrames by GDP per capita in ascending order

df_2020_sorted = df_2020.sort_values(by='gdp', ascending=True)

df_2021_sorted = df_2021.sort_values(by='gdp', ascending=True)

df_2022_sorted = df_2022.sort_values(by='gdp', ascending=True)


# Get the country with the lowest GDP per capita for each year

lowest_gdp_2020 = df_2020_sorted.iloc[:5]


lowest_gdp_2020

lowest_gdp_2021 = df_2021_sorted.iloc[:5]

lowest_gdp_2021

lowest_gdp_2022 = df_2022_sorted.iloc[:5]

lowest_gdp_2022

# Data for the top 5 countries with the highest GDP in 2020

countries_2020 = ['Luxembourg', 'Ireland', 'Singapore', 'Switzerland', 'Norway']

gdp_per_capita_2020 = [111751.312500, 91356.851562, 82886.781250, 67765.882812, 63776.160156]


# Data for the top 5 countries with the highest GDP in 2021

countries_2021 = ['Luxembourg', 'Ireland', 'Singapore', 'Switzerland', 'Norway']
```

gdp_per_capita_2021 = [115683.492188, 102785.492188, 82886.781250, 70097.367188, 65909.000000]

Data for the top 5 countries with the highest GDP in 2022

countries_2022 = ['Ireland', 'Singapore', 'United States', 'Denmark', 'United Kingdom']

gdp_per_capita_2022 = [113870.789062, 82886.781250, 64702.980469, 59704.230469, 44949.093750]

Data for the top 5 countries with the lowest GDP in 2020

countries_2020_lowest = ['Greenland', 'Mozambique', 'Mali', 'Uganda', 'Kenya']

gdp_per_capita_2020_lowest = [324.000000, 1233.425049, 2123.828125, 2240.490234, 4497.362793]

Data for the top 5 countries with the lowest GDP in 2021

countries_2021_lowest = ['Greenland', 'Central African Republic', 'Kyrgyzstan', 'Kenya', 'Honduras']

gdp_per_capita_2021_lowest = [324.000000, 837.504700, 4726.196777, 4745.637207, 5572.176758]

Data for the top 5 countries with the lowest GDP in 2022

countries_2022_lowest = ['Greenland', 'Bhutan', 'Kyrgyzstan', 'Bangladesh', 'El Salvador']

gdp_per_capita_2022_lowest = [324.000000, 2977.000000, 4726.196777, 6263.004883, 9397.530273]

Creating the line chart

plt.figure(figsize=(15, 8))

Plotting for highest GDP

plt.plot(countries_2020, gdp_per_capita_2020, marker='o', color='b', linestyle='-', label='2020')

plt.plot(countries_2021, gdp_per_capita_2021, marker='o', color='r', linestyle='-', label='2021')

plt.plot(countries_2022, gdp_per_capita_2022, marker='o', color='g', linestyle='-', label='2022')

Plotting for lowest GDP

plt.plot(countries_2020_lowest, gdp_per_capita_2020_lowest, marker='x', color='b', linestyle='--', label='2020 (Lowest GDP)')

plt.plot(countries_2021_lowest, gdp_per_capita_2021_lowest, marker='x', color='r', linestyle='--', label='2021 (Lowest GDP)')

plt.plot(countries_2022_lowest, gdp_per_capita_2022_lowest, marker='x', color='g', linestyle='--', label='2022 (Lowest GDP)')

Adding title and labels

plt.title('GDP per Capita for Top and Bottom 5 Countries (2020-2022)')

plt.xlabel('Country')

plt.ylabel('GDP per Capita')

Rotating x-axis labels for better readability

plt.xticks(rotation=45)

Adding legend

plt.legend()

Display the chart

plt.grid(True)

plt.tight_layout()

```
plt.show()

reg_2020 = ['Europe and Central Asia',  'East Asia and the Pacific']

gdp_per_capita_2020_1 = [111751.312500, 82886.781250]



reg_2021 = ['Europe and Central Asia', 'East Asia and the Pacific',  ]

gdp_per_capita_2021_1= [115683.492188, 102785.492188]



reg_2022 = ['Europe and Central Asia', 'East Asia and the Pacific', 'North America', ]

gdp_per_capita_2022_1= [113870.789062, 82886.781250, 64702.980469]


reg_2020_lowest = ['Sub-Saharan Africa']

gdp_per_capita_2020_lowest_1 = [1233.425049]


reg_2021_lowest = ['Latin America and the Caribbean']

gdp_per_capita_2021_lowest_1 = [5572.176758]

reg_2022_lowest = ['South Asia']

gdp_per_capita_2022_lowest_1 = [2977.000000]


# Creating the line chart

plt.figure(figsize=(10, 6))


# Plotting for 2020

plt.plot(reg_2020, gdp_per_capita_2020_1, marker='o', color='b', linestyle='-', label='2020')

plt.plot(reg_2021, gdp_per_capita_2021_1, marker='o', color='r', linestyle='-', label='2021')

plt.plot(reg_2022, gdp_per_capita_2022_1, marker='o', color='g', linestyle='-', label='2022')


plt.plot(reg_2020_lowest, gdp_per_capita_2020_lowest_1, marker='x', color='b', linestyle='--', label='2020 (Lowest GDP)')

plt.plot(reg_2021_lowest, gdp_per_capita_2021_lowest_1, marker='x', color='r', linestyle='--', label='2021 (Lowest GDP)')

plt.plot(reg_2022_lowest, gdp_per_capita_2022_lowest_1, marker='x', color='g', linestyle='--', label='2022 (Lowest GDP)')


# Adding title and labels

plt.title('GDP per Capita Reg with Highest GDP (2020-2022)')

plt.xlabel('Region')

plt.ylabel('GDP per Capita')


# Rotating x-axis labels for better readability

plt.xticks(rotation=45)


# Adding legend

plt.legend()


# Display the chart

plt.grid(True)

plt.tight_layout()

plt.show()

# Sort the DataFrames by Gini coefficient in descending order

df_2020_sorted = df_2020.sort_values(by='gini', ascending=False)

df_2021_sorted = df_2021.sort_values(by='gini', ascending=False)

df_2022_sorted = df_2022.sort_values(by='gini', ascending=False)
```

Get the country with the highest Gini coefficient for each year

highest_gini_2020 = df_2020_sorted.iloc[:5]

highest_gini_2021 = df_2021_sorted.iloc[:5]

highest_gini_2022 = df_2022_sorted.iloc[:5]

highest_gini_2020

highest_gini_2021

highest_gini_2022

Sort the DataFrames by Gini coefficient in descending order

df_2020_sorted = df_2020.sort_values(by='gini', ascending=True)

df_2021_sorted = df_2021.sort_values(by='gini', ascending=True)

df_2022_sorted = df_2022.sort_values(by='gini', ascending=True)

Get the country with the highest Gini coefficient for each year

lowest_gini_2020 = df_2020_sorted.iloc[:5]

lowest_gini_2021 = df_2021_sorted.iloc[:5]

lowest_gini_2022 = df_2022_sorted.iloc[:5]

lowest_gini_2020

lowest_gini_2021

lowest_gini_2022