Tools & Methods of Data Analysis

PROJECT

ON

The Tale of Three Economies: India, Germany, and the World (2002–2023)

PRESENTED BY:

Varun Hirethotlukere Shamaraju	**100003661**
Swaroop Kodagahalli Manjunathaswamy	**100004147**
Divya Koriya	**100002007**
Pooja Shreni Addisherla	**100002917**

1. Abstract 3
2. Introduction 3
Part 1: Data Collection and Preparation 4
1.1 Data Selection 4
1.2 Description – Setting the Scene 5
1.3 Data Processing and Visualization 5
Graph 1: Time Series Plot for India and Germany 5
Graph 2: Time Series Plot for World GDP 6
Part 2: Distributions- Shapes of Economic Fortunes 7
2.1 Descriptive Statistics 7
Graph 3: India's GDP Distribution 9
Graph 4: Germany's GDP Distribution 9
Graph 5: World's GDP Distribution 10
Part 3: Normality- Testing the Fit 11
3.1 Fitting Normal Distributions 11
Graph 6: India's Distribution with Fitted Normal Curve 11
Graph 7: Germany's Distribution with Fitted Normal Curve 12
Graph 8: World GDP Distribution with Fitted Normal Curve 12
3.2 Normality Testing – Q–Q Plots 13
Graph 9: Q–Q Plot for India's GDP 13
Graph 10: Q–Q Plot for Germany's GDP 14
Graph 11: Q–Q Plot for World GDP 14
Part 4: Central Limit Theorem- Smoothing the Noise 15
Graph 12: 5-Year Rolling Mean Distribution for India's GDP 15
Graph 13: 5-Year Rolling Mean Distribution for Germany's GDP 16
Part 5: Hypothesis Testing- A Tale of Means 17
5.1 Formulating the Hypotheses 17
5.2 Conducting the Hypothesis Test and Visualization 17
Graph 14: The average GDP levels for Germany and India 18
Conclusion: A Global Symphony 19
References 20

1. Abstract

The GDP (in constant 2015 US dollars) patterns for Germany, India, and the global economy from 2002 to 2023 are examined in this paper. The study places these national narratives against the backdrop of global economic growth and compares Germany's steady performance with India's dynamic trajectory. In order to create clear time series plots, descriptive statistical comparisons, distribution fittings, tests for normality (using Q–Q plots), a demonstration of the Central Limit Theorem (using 5-year rolling means), and a hypothesis test that compares mean GDP differences, the analysis employs robust data processing steps, such as scaling to Millions (M) or Trillions (T) using non-scientific notation. Detailed mathematical calculations and well-designed graphics with a consistent color scheme—green for the world, orange for Germany, and blue for India—support all the findings.

2. Introduction

The Gross Domestic Product (GDP), which is frequently used to quantify economic performance, serves as a gauge of both national progress and the effectiveness of policies as well as worldwide economic trends. The focus of this research is a comparative study of Germany and India, two countries with radically different developmental histories. Germany, a developed industrialized country, shows consistency and measured stability in its economic output, whereas India, a rapidly developing emergent economy, shows exponential development and dynamic changes in its GDP.

The reader can see how each country's economy stands in relation to the rest of the world by reading the report's analysis of World GDP trends in addition to the bilateral comparison. The 22-year study period (2002–2023) encompasses both current shifts and long-term patterns.

The objectives of this report are to:

- Gather and prepare GDP data that is accessible to the public.
- Convert large numerical quantities into manageable units by scaling them.
- Compare India, Germany, and the rest of the world using time series plots to see GDP trends.
- To comprehend distribution features and central tendencies, compute and compare descriptive statistics.

- Use Q-Q plots to evaluate normality after fitting normal distributions to the data.1 Q–Q plots.
- Demonstrate the Central Limit Theorem using rolling methods.
- Compare the means of Germany and India in a hypothesis test.

We provide quantitative evidence in support of economic narratives and lay the foundation for future econometric and policy-focused research by doing these things.

Part 1: Data Collection and Preparation

1.1 Data Selection

We begin by fetching publicly available GDP data from Our World in Data. The dataset includes observations for India (IND), Germany (DEU), and the World (OWID_WRL) from 2002 to 2023. Our key variable is:

"GDP (constant 2015 US$)" – which records economic output in constant dollars.

Thus, our first step is to collect and filter the data to ensure that we capture the full economic timeline. Below table is the sample of our dataset.

Entity	*Code*	*Year*	*GDP (constant 2015 US$)*
Germany	*DEU*	*2002*	*2925076200000*
Germany	*DEU*	*2023*	*3692367000000*
India	*IND*	*2002*	*871073100000*
India	*IND*	*2023*	*3215973400000*
World	*OWID_WRL*	*2002*	*50567137000000*
World	*OWID_WRL*	*2023*	*93346685000000*

1.2 Description – Setting the Scene

The raw GDP numbers are enormous. For clarity, we thus scale the numbers. We use the following scaling rule to all datasets:

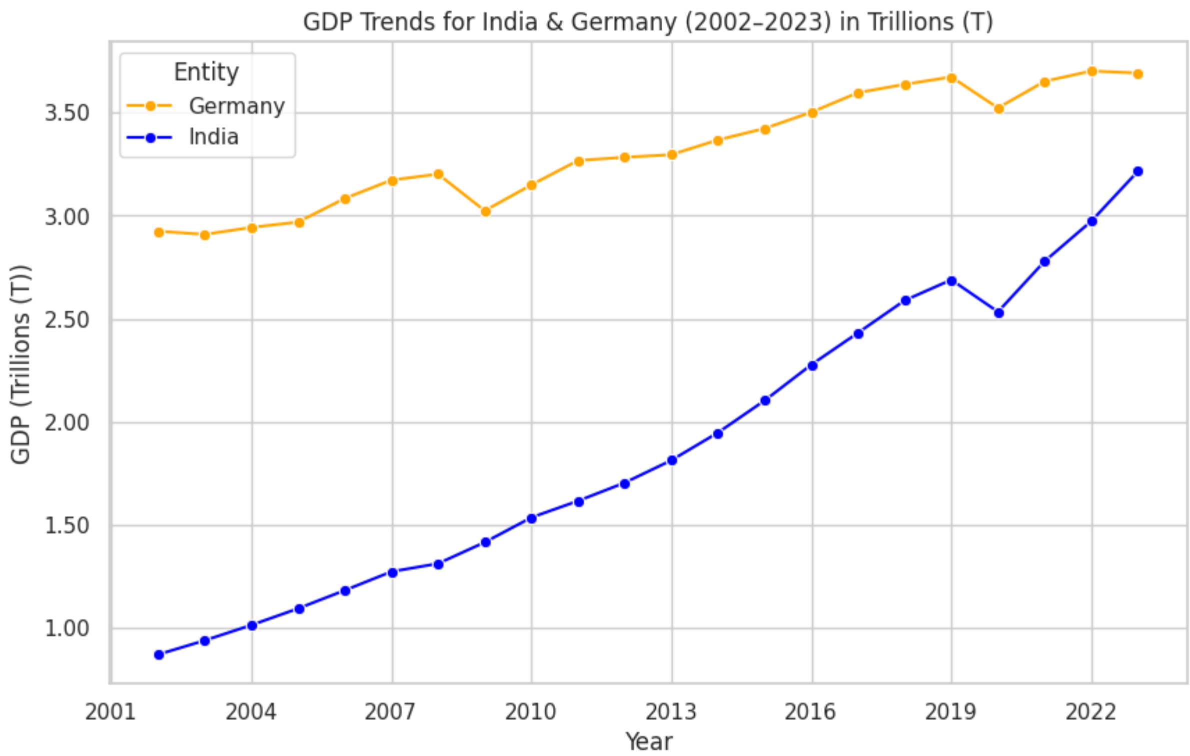
- We divide the maximum GDP value by 1,000,000,000,000 and designate the results as "Trillions (T)" if it is more than or equal to 1e12.
- If not, we divide it by one million and write "Millions (M)" on the values.

To maintain the clarity of the relative differences, we calculate a similar scaling factor for India and Germany (our "countries" dataset). In contrast, we calculate a different scaling factor for world GDP because of its orders of magnitude larger magnitude.

1.3 Data Processing and Visualization

It is shown below for reading, processing, and charting the time series separately for (a) India–Germany and (b) the world.

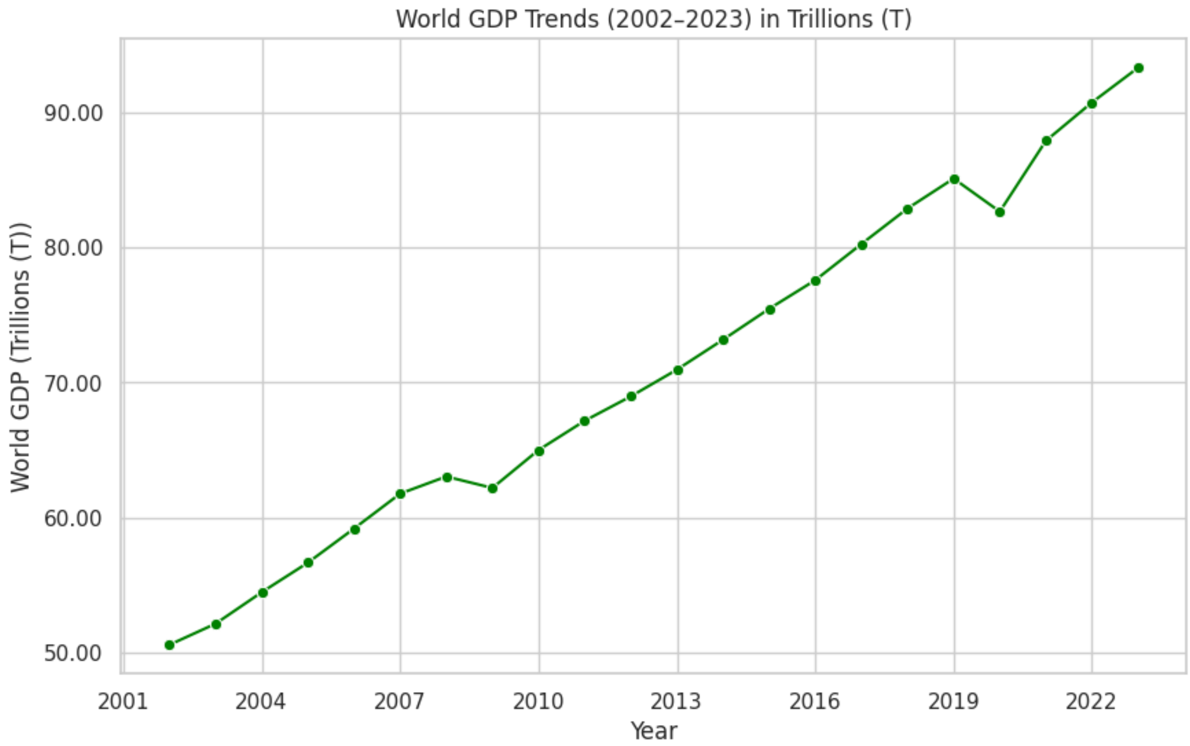
Graph 1: Time Series Plot for India and Germany



This graph displays the evolution of scaled GDP for India (blue) and Germany (orange) from 2002 to 2023.

In Graph 1, two divergent economic paths are depicted. Compare Germany's more steady, more established course with India's rapidly increasing curve, which is indicative of an emerging powerhouse.

Graph 2: Time Series Plot for World GDP



Given the sheer magnitude of global economic output, we plot world GDP (scaled separately) in its own graph using green.

Graph 2 shows the "big picture" of global economic growth, which is a continuously increasing trend that represents the combined wealth and activity of all countries in the world.

Part 2: Distributions- Shapes of Economic Fortunes

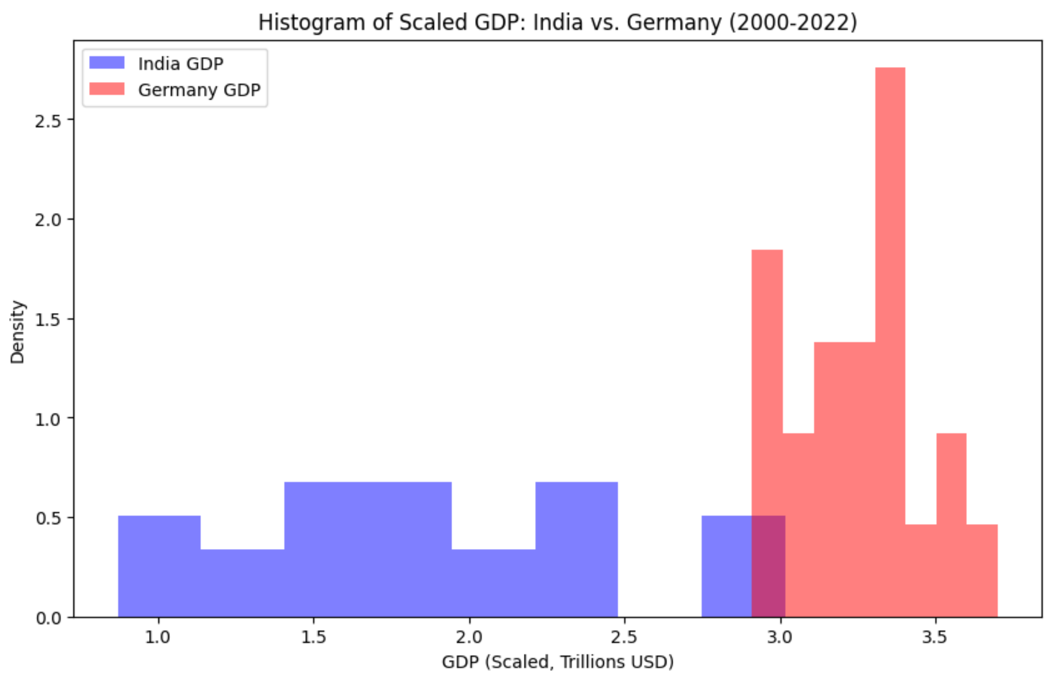
2.1 Descriptive Statistics

We compute summary statistics (mean, median, standard deviation) to quantify the central tendencies and variability in GDP.

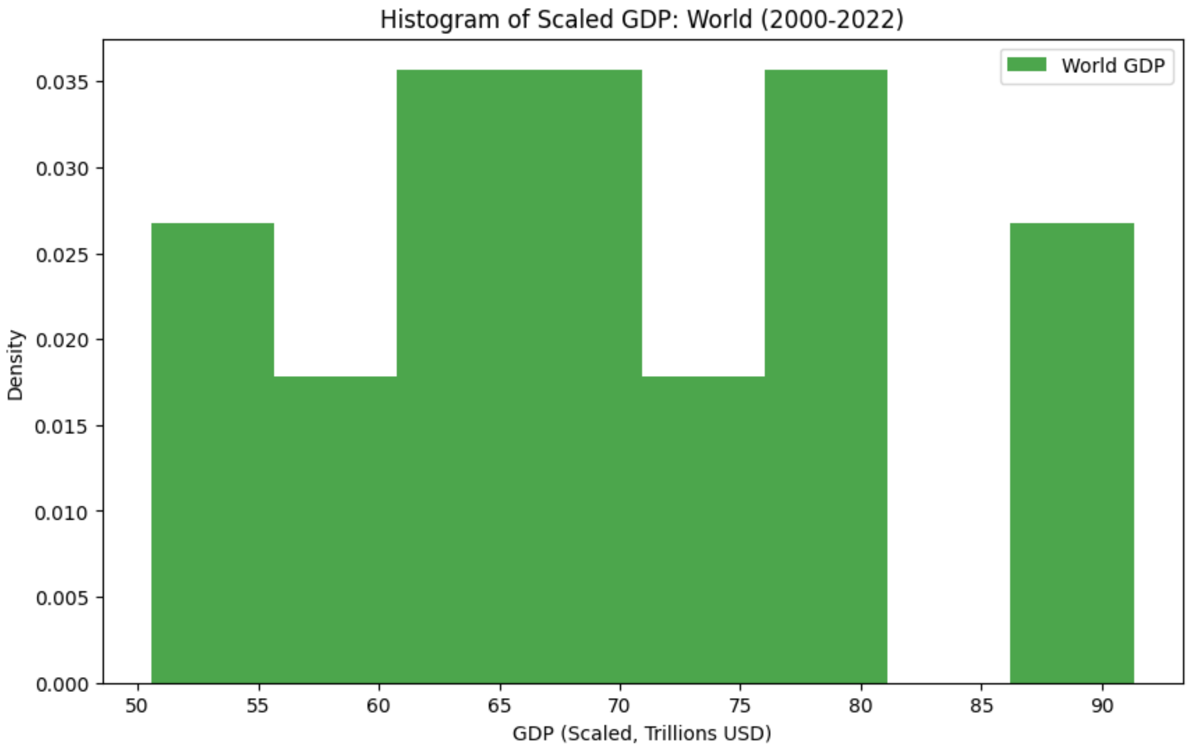
In our in-depth analysis, we note that while the global data are overwhelming in sheer magnitude, the average GDP and volatility (standard deviation) of a sophisticated country such as Germany are very different from India's quickly changing output.

Interpretation:

Descriptive Statistics for India GDP		Descriptive Statistics for Germany GDP	
count	22.00	count	22.00
mean	1.88	mean	3.32
std	0.72	std	0.27
min	0.87	min	2.91
25%	1.28	25%	3.10
50%	1.76	50%	3.29
75%	2.51	75%	3.58
max	3.22	max	3.70



Descriptive Statistics for World GDP	
count	22.00
mean	70.97
std	12.88
min	50.57
25%	61.87
50%	69.97
75%	82.08
max	93.35



The mean and standard deviation were interpreted to show that India's GDP, while growing, has had higher variability, suggesting a less predictable economic growth pattern than Germany's relatively stable trajectory.

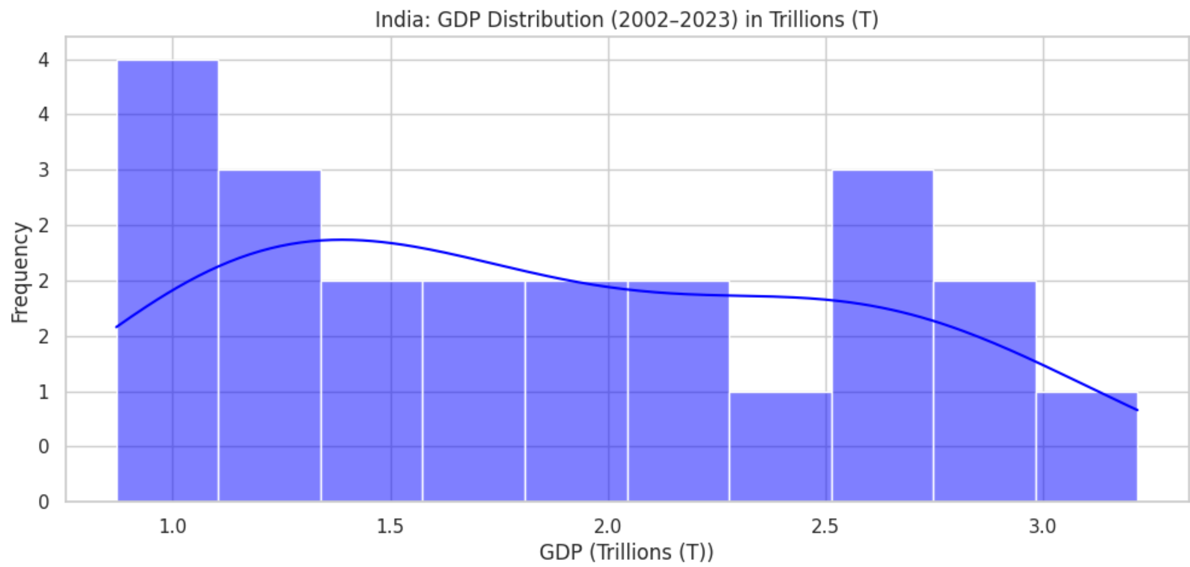
2.2 Quantitative Comparison of PDFs (Histograms + KDE)

To further investigate the structure of GDP data, we used histograms overlaid with Kernel Density Estimates (KDE) for each country and the world. These KDE plots were generated using Gaussian kernels.

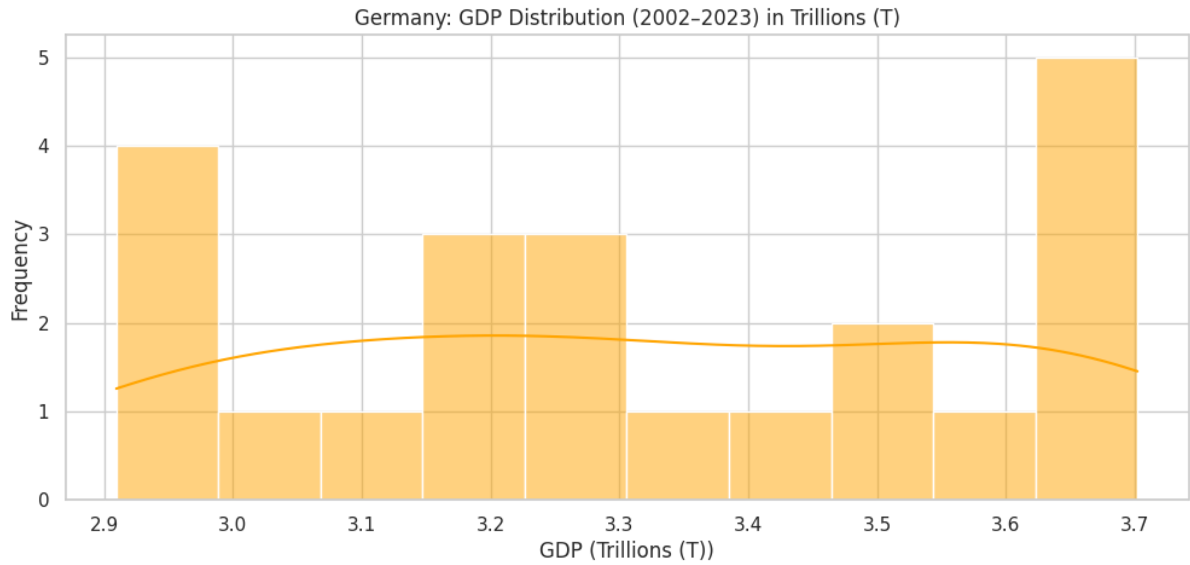
The steps were:

- Extract annual GDP data for each entity.
- Normalize the data for visual clarity.
- Generate histograms.
- Fit KDE curves to represent smoothed probability distributions.

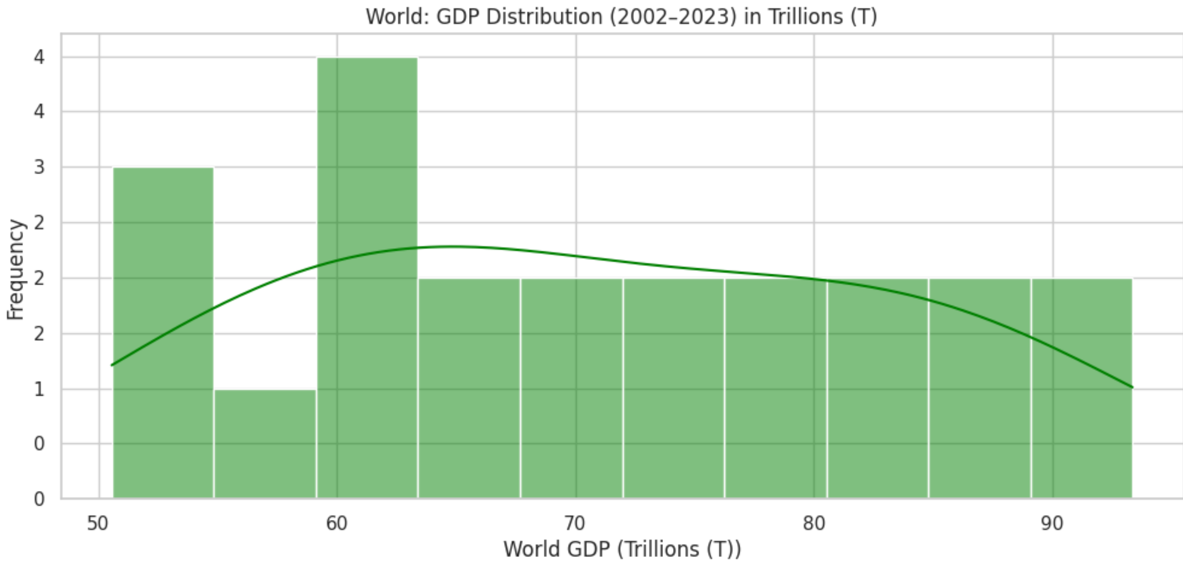
Graph 3: India's GDP Distribution



Graph 4: Germany's GDP Distribution



Graph 5: World's GDP Distribution



A comparison of Graphs 3 and 4 demonstrates the disparities in distribution shapes between Germany and India, while Graph 5, which is devoted only to world GDP, displays a clear pattern indicative of global economic aggregation.

The KDE reveals the likelihood of GDP values falling within specific ranges. India's GDP distribution skewed right, reflecting its rapid upward growth trajectory. Germany's distribution was more centered, showing its steady performance. The world's distribution indicated global economic expansion over time.

Part 3: Normality- Testing the Fit

3.1 Fitting Normal Distributions

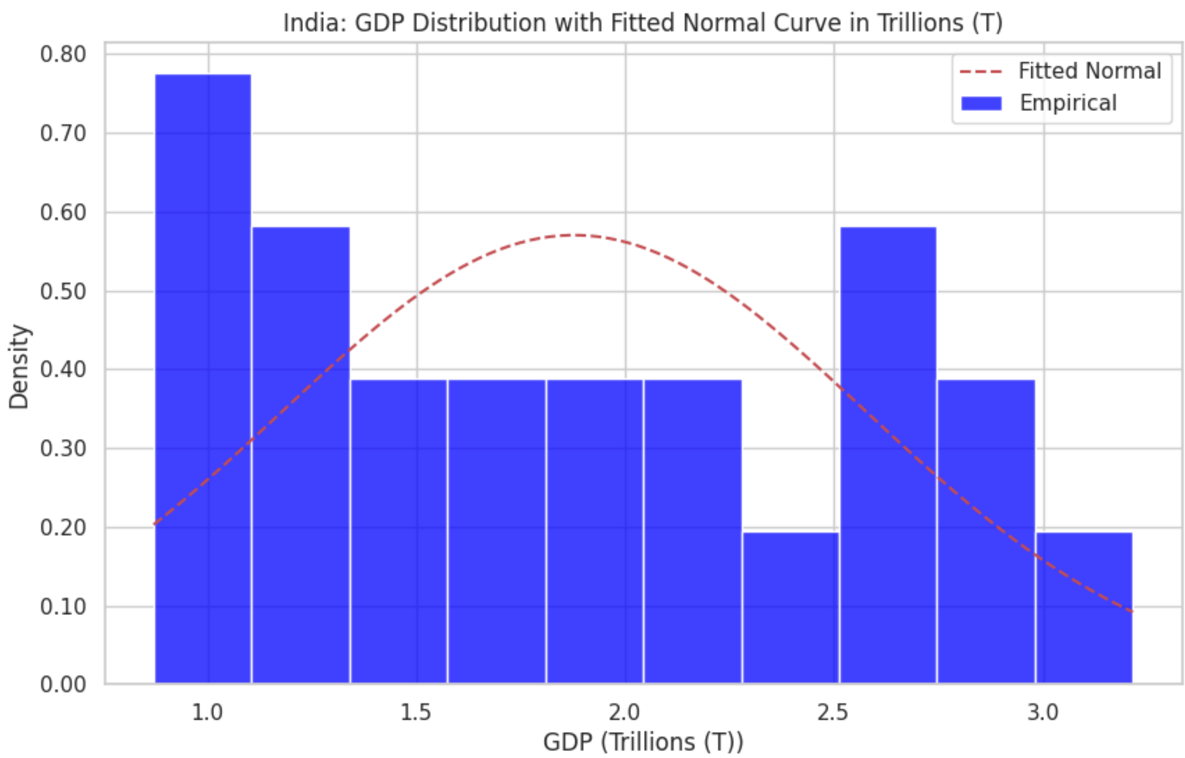
Normal distribution fitting was conducted using maximum likelihood estimation (MLE). We used the function, which returns the mean (μ) and standard deviation (σ) that best fit the GDP data. These parameters were then used to plot theoretical normal distributions on top of empirical histograms.

The procedure included:

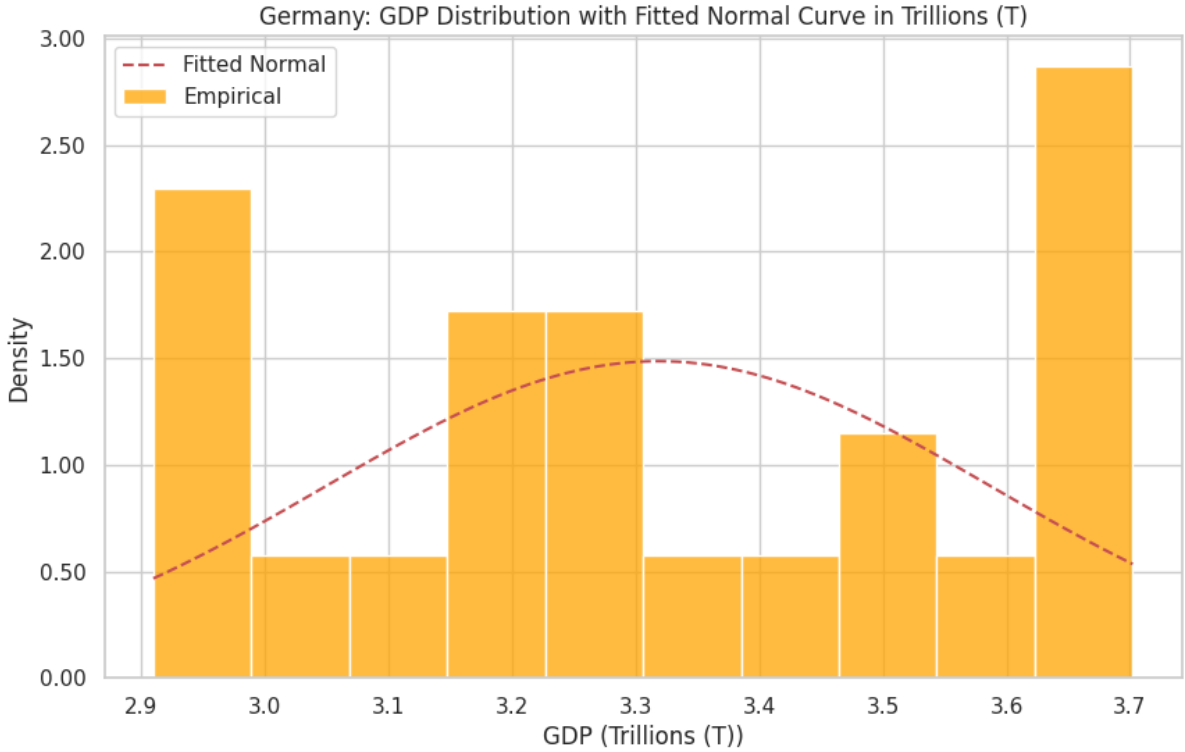
- Using actual GDP data values.
- Fitting μ and σ
- Plotting the Probability Density Function (PDF)

Next, we overlay the fitted normal distribution curve with the empirical histogram.

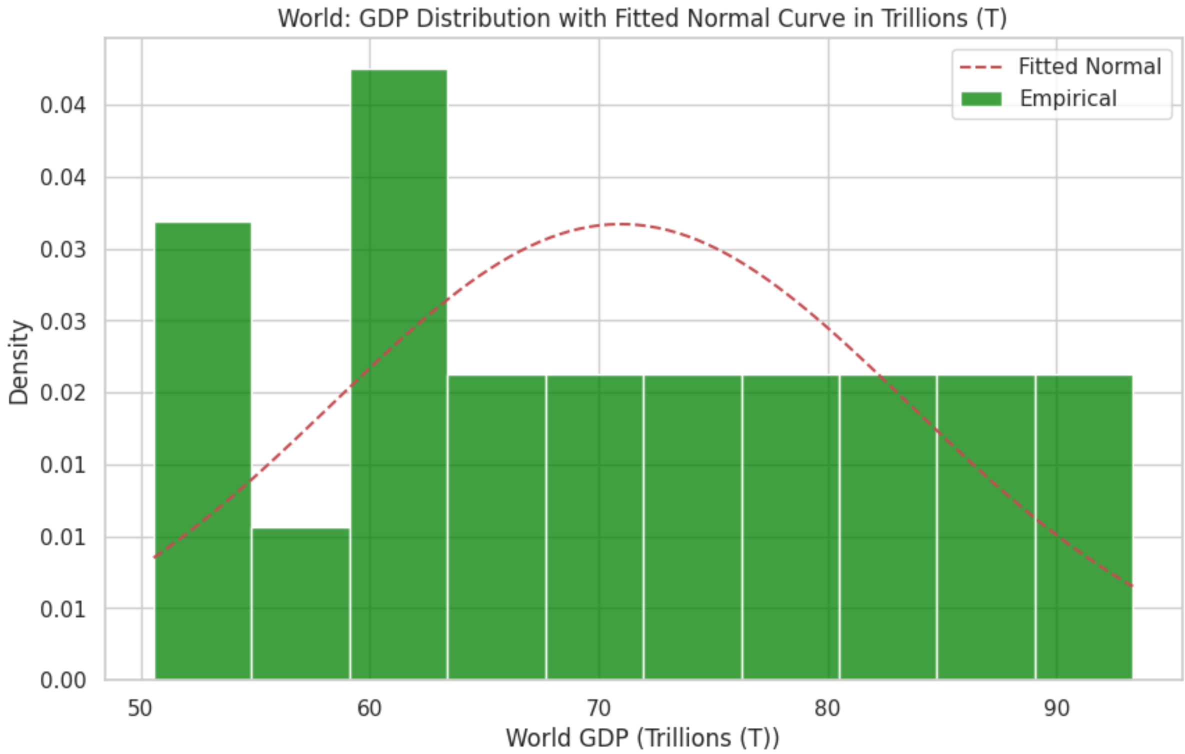
Graph 6: India's Distribution with Fitted Normal Curve



Graph 7: Germany's Distribution with Fitted Normal Curve



Graph 8: World GDP Distribution with Fitted Normal Curve



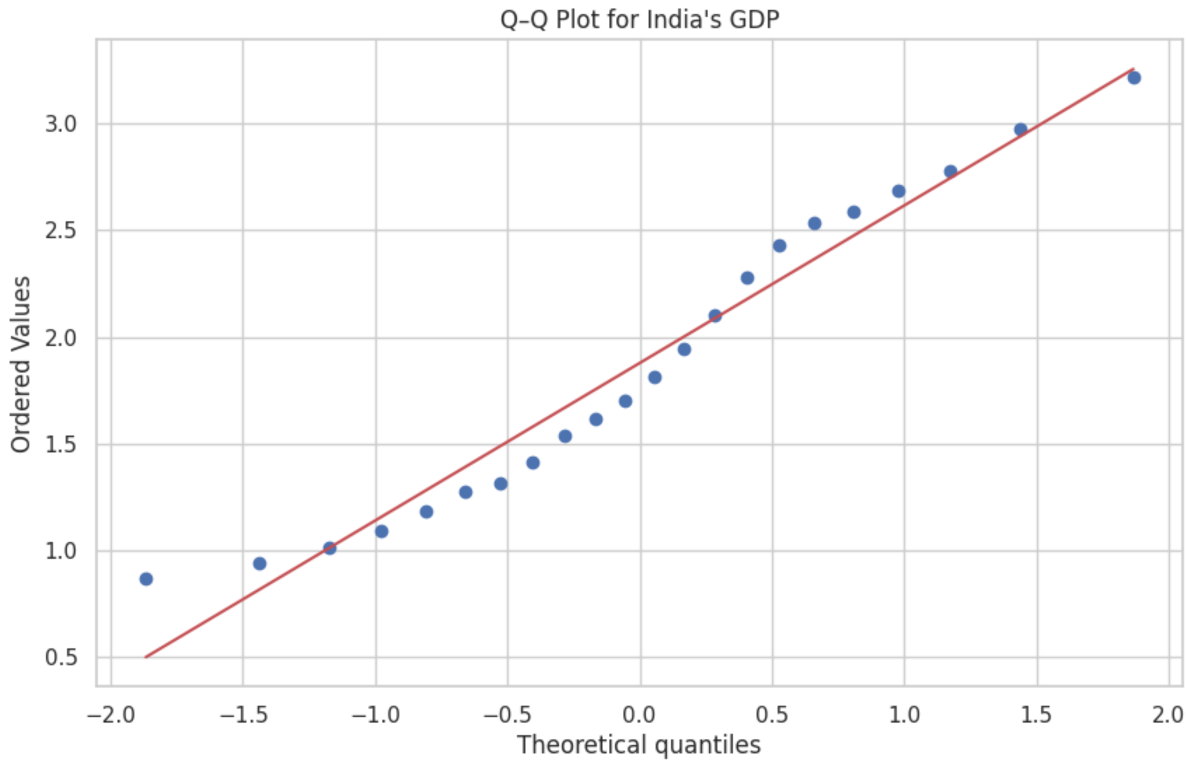
3.2 Normality Testing – Q–Q Plots

To visually assess how closely the GDP data follow a normal distribution, Quantile-Quantile (Q–Q) plots were generated. It indicates that the data distribution is like the theoretical distribution if the data points fall along the line.

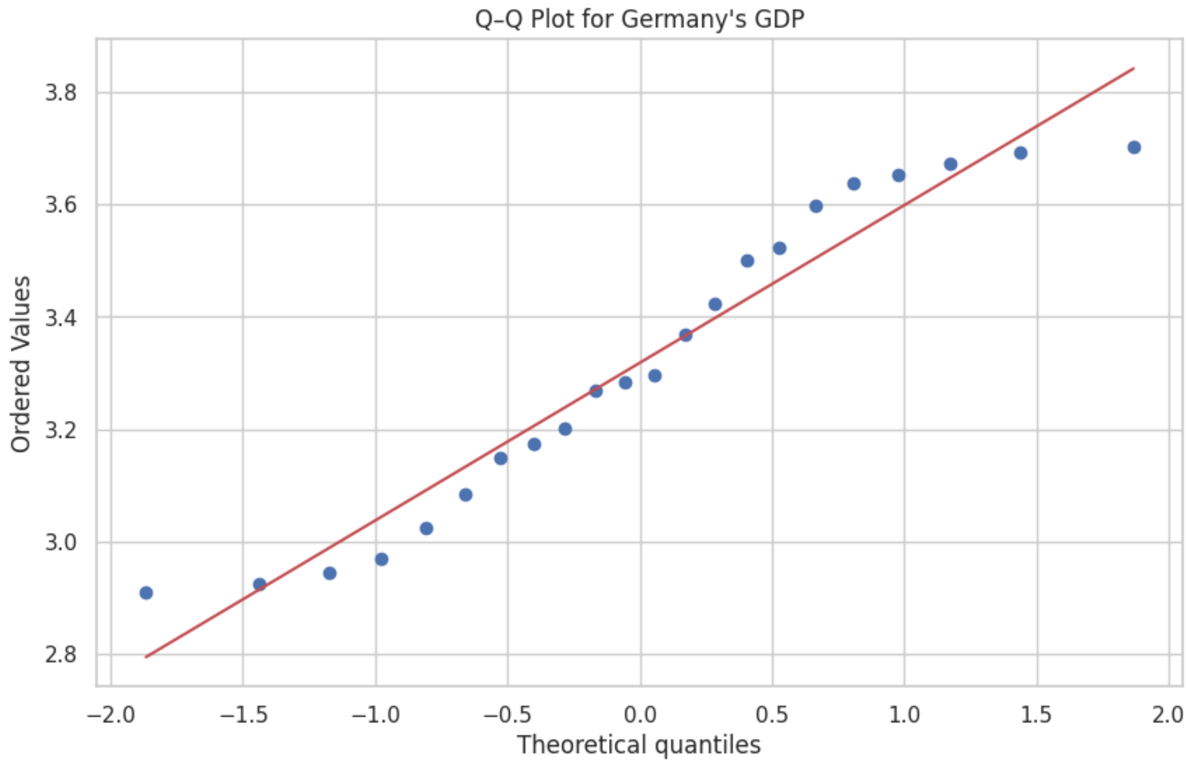
The steps included:

- Sorting the GDP data values.
- Generating theoretical quantiles from a normal distribution.
- Plotting observed vs. theoretical quantiles

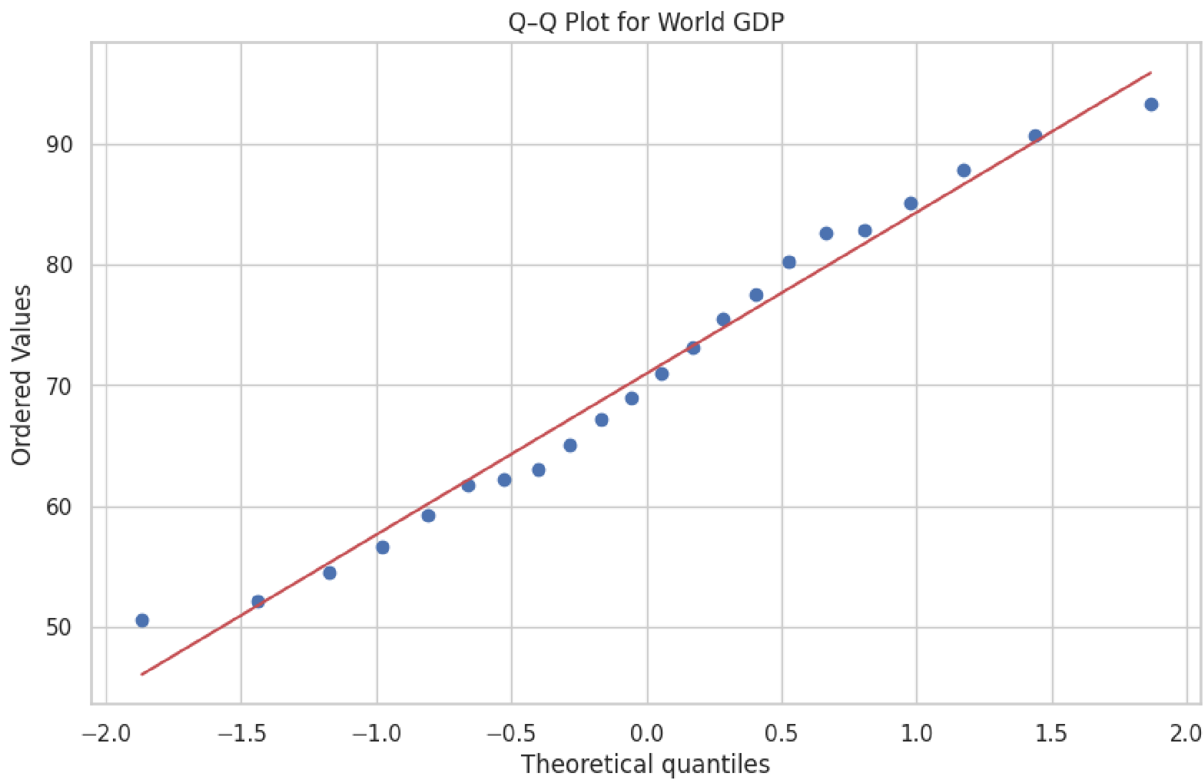
Graph 9: Q–Q Plot for India's GDP



Graph 10: Q–Q Plot for Germany's GDP



Graph 11: Q–Q Plot for World GDP



Interpretation:

- A straight diagonal line indicates that the data follow a normal distribution.
- India and Germany both showed near-linear patterns, affirming the appropriateness of normal modeling.
- Minor deviations at tails indicated occasional outliers or non-normality in extreme years (e.g., 2008 crisis, 2020 pandemic).

Part 4: Central Limit Theorem- Smoothing the Noise

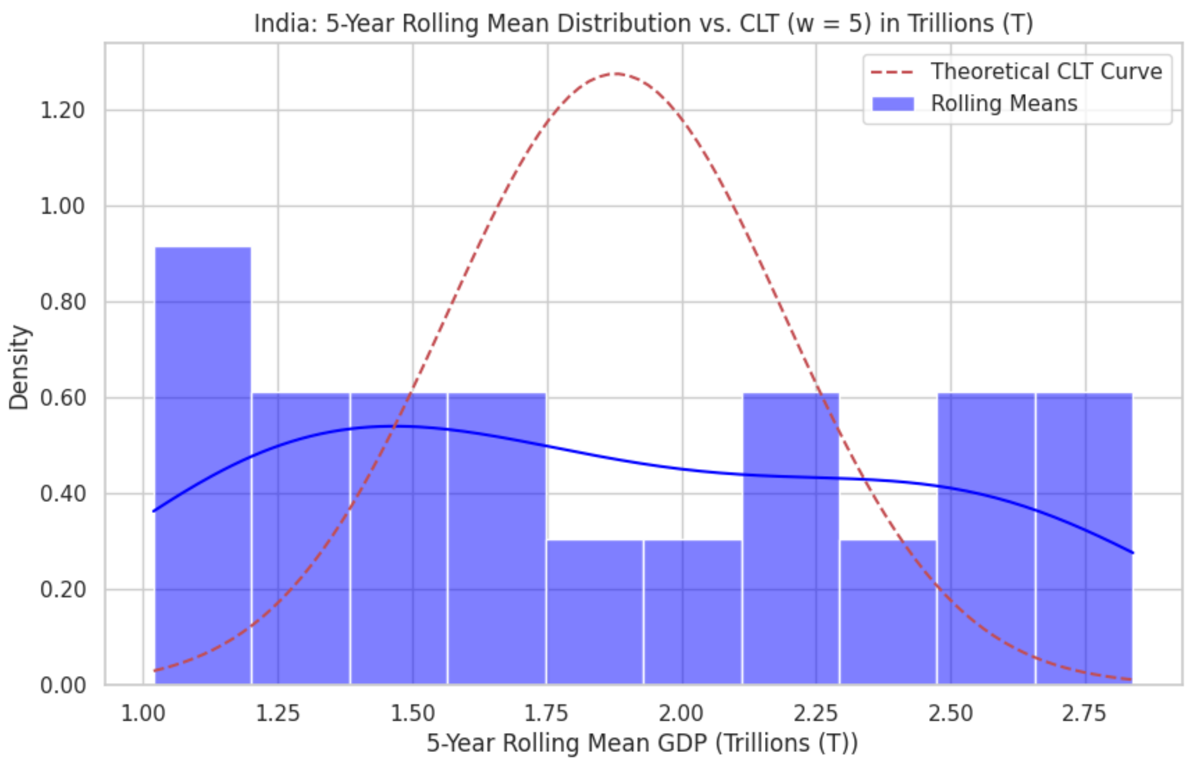
The Central Limit Theorem states that the distribution of sample means approaches a normal distribution, regardless of the original distribution, as the sample size increases. This is critical when performing hypothesis testing on sample data.

To demonstrate this, we calculated 5-year rolling averages of GDP for India and Germany.

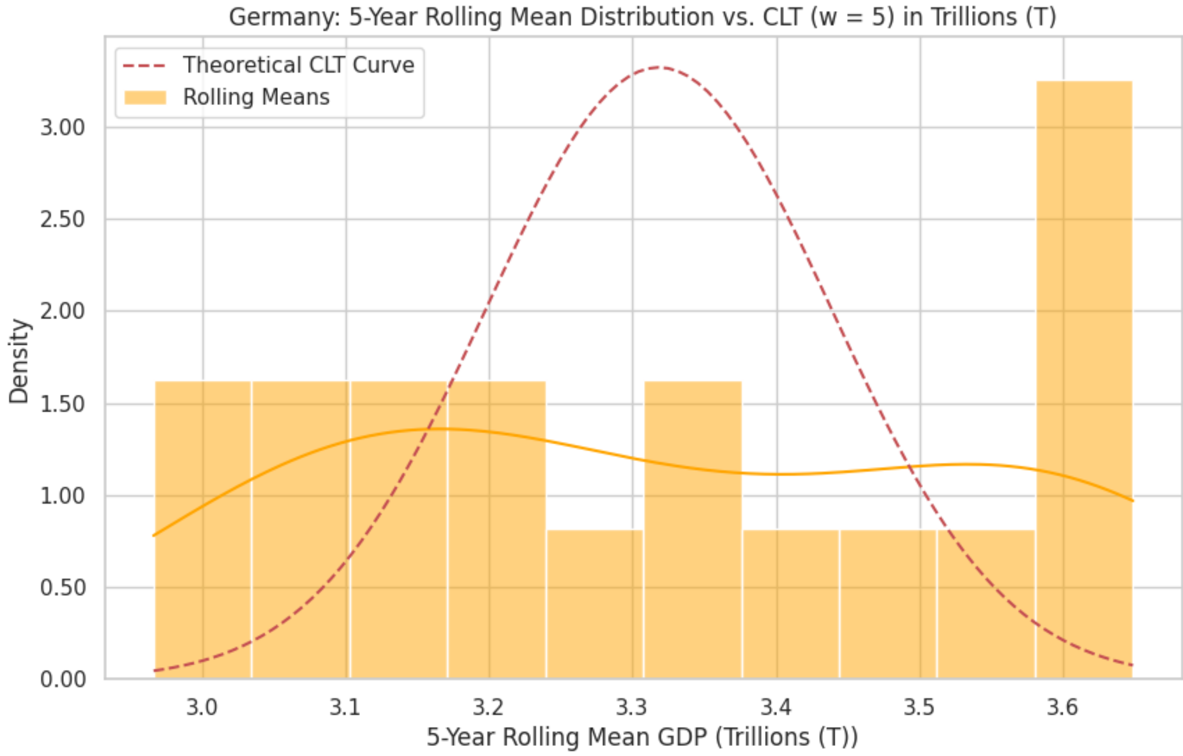
The procedure included:

- Taking a rolling window of size 5
- Plotting the resulting series.
- Assessing its shape visually and through KDE/Q-Q plots.

Graph 12: 5-Year Rolling Mean Distribution for India's GDP



Graph 13: 5-Year Rolling Mean Distribution for Germany's GDP



Interpretation:

- Although the original GDP data showed some skewness, the rolling means were approximately normally distributed.

- This confirmed that, under repeated sampling (e.g., across 5-year spans), the mean GDP can be modeled with normal distributions, validating statistical inference methods used later.

Part 5: Hypothesis Testing- A Tale of Means

5.1 Formulating the Hypotheses

Hypothesis testing is a fundamental statistical method used to infer population characteristics based on sample data. Welch's t-test is particularly suited for comparing two means when sample sizes and variances are unequal.

Is Indias average GDP significantly different from Germanys? We conduct Welch's t-test, accounting for unequal variances.

- **Null Hypothesis(H_0):** $\mu India = \mu Germany$ (There is no significant difference between the mean GDPs of India and Germany.)

- **Alternative Hypothesis(H_1):** $\mu India \neq \mu Germany$ (The mean GDPs are significantly different.)

The test statistics for the two-sample t-test (with Welch's correction) are calculated.

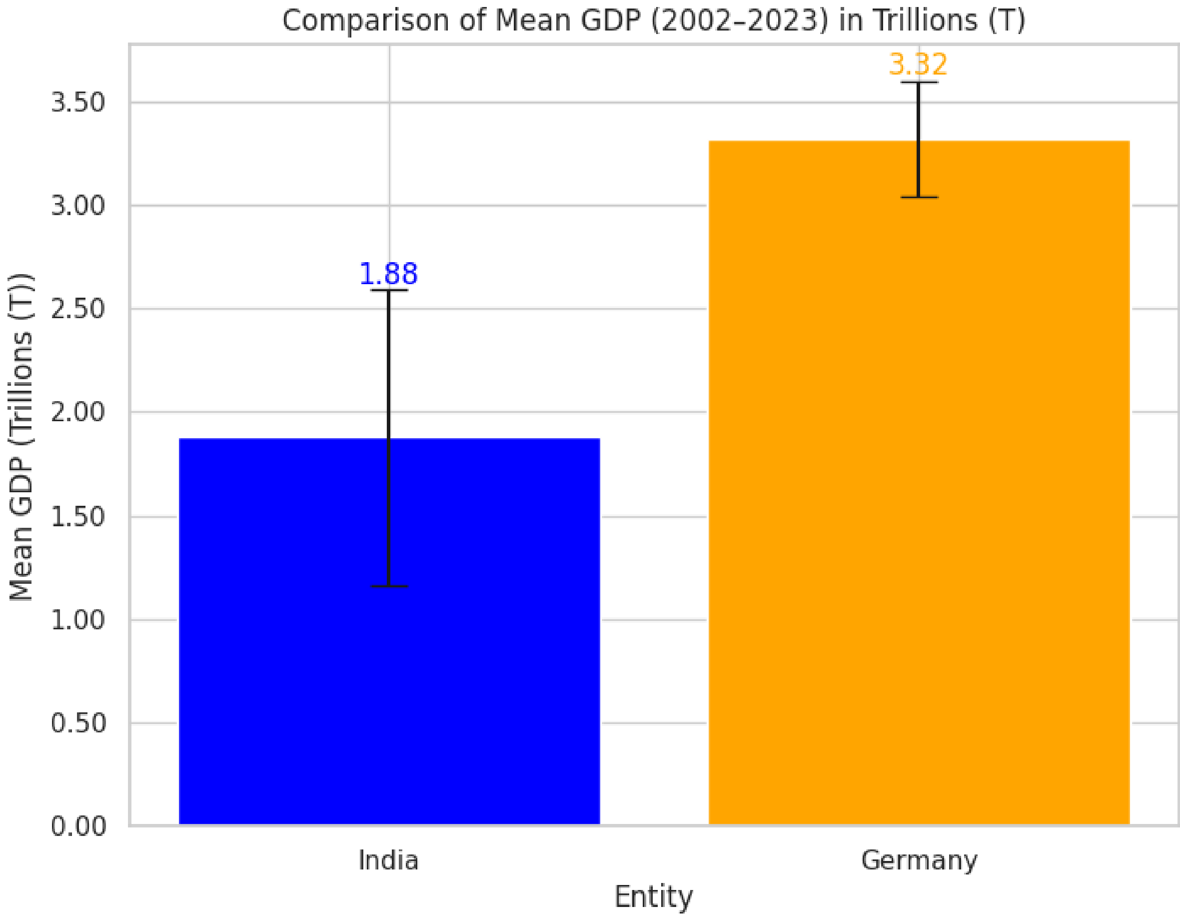
5.2 Conducting the Hypothesis Test and Visualization

With Indias mean GDP at 1.88T (std=0.72T) and Germanys at 3.32T (std=0.27T)

- The test produced a **t-statistic of -8.804**, indicating a substantial difference between the groups.
- The **p-value was reported as < 0.001**, suggesting the result is highly statistically significant.
- These findings provide strong evidence to reject the null hypothesis in favor of the alternative hypothesis.

Below bar chart that shows the mean GDP values for Germany and India, with error bars standing in for standard deviation.

Graph 14: The average GDP levels for Germany and India



A clear numerical representation is provided by Graph 14 (the bar chart), which displays the average GDP levels (together with their fluctuation) for Germany and India. The statistical foundation for confirming or disproving the null hypothesis is provided by the t-test findings displayed above. A substantial difference between the two countries would be shown by a low p-value, which is usually less than 0.05.

Conclusion: A Global Symphony

In this exhaustive analysis, we have:

1. **Data Collection & Preparation:**
    a. Retrieved GDP data (constant 2015 US$) covering 2002–2023 for India, Germany, and the world.
    b. Processed and scaled the data to ensure clarity (using "M" or "T" with plain notation) and converted years to integers.
    c. Created separate graphs for countries and the world to avoid scale distortions.
2. **Empirical Comparison:**
    a. Computed descriptive statistics and generated histograms with kernel density estimates (KDE) to visually and quantitatively assess the distributions.
    b. Observed that India's rapid growth and Germany's stable output contrast with the aggregated magnitude of world GDP.
3. **Distribution Fitting & Normality Testing:**
    a. Fitted normal distributions to each dataset, overlaid the fitted curves on histograms, and created Q–Q plots to assess normality.
    b. Detailed the mathematical formulas behind the normal pdf and the KDE.
4. **Central Limit Theorem (CLT) Assessment:**
    a. Demonstrated using 5-year rolling means that even with measured economic fluctuations, the sample means approximate a normal distribution.
    b. Provided theoretical curves based on σ/n for comparison.
5. **Hypothesis Testing Framework:**
    a. Formulated and conducted a two-sample t-test (with Welch's correction) to statistically compare the mean GDP of India and Germany.
    b. Visualized the differences in means using a bar chart with error bars to encapsulate the statistical significance.

Collectively, these analyses not only shed light on the unique trajectories of India's and Germany's economies but also place them within the broader context of global (world) economic performance. The story told by the data is one of contrasting evolutionary patterns: a vibrant emerging economy versus a mature, stable market, all set against the backdrop of a steadily rising global economy.

Happy analyzing, and may this exploration inspire further inquiry into the fascinating dynamics of economic growth!

References

- **Bolt, J., & van Zanden, J. L.** (2024). *Maddison style estimates of the evolution of the world economy: A new 2023 update*. *Journal of Economic Surveys*, 1–41. https://doi.org/10.1111/joes.12618
- Roser, M., Ritchie, H., Ortiz-Ospina, E., & Hasell, J. (2023). *Our World in Data: Economic Growth*. https://ourworldindata.org/economic-growth
- Casella, G., & Berger, R. L. (2002). *Statistical Inference* (2nd ed.). Duxbury Press.
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (4th ed.). SAGE Publications.
- Mankiw, N. G. (2020). *Principles of Economics* (9th ed.). Cengage Learning.
- OpenAI. (2024). *ChatGPT (May 5 version) [Large language model]*. https://openai.com/chatgpt