***ANALYSIS OF GDP VARIATION OF VARIOUS COUNTRIES FROM 2000-2017***

Parv Nangalia

PES12012700249

[nangaliaparv@gmail.com](mailto:nangaliaparv@gmail.com)

R. Ananth

PES1201700286

[rananth@gmail.com](mailto:rananth@gmail.com)

Vishwas. N. S

PES1201701321

[vishwas3151999@gmail.com](mailto:vishwas3151999@gmail.com)

Section 3G

DEPARTMENT OF CSE

PES UNIVERSITY

***ABSTRACT*:-**

We as a group have conducted statistical analysis on the trends of GDP for various country’s around the world. The GDP values for every country is made out with a common measuring unit, i.e., the present US dollar conversion rates. We have analysed the rise and fall of GDP values of various countries over a period of 18 years (2000 to 2017) and infer from it about countries which underwent rapid development and countries which are rather stable in terms of production. With the given data we also hypothesize if different countries developed together during particular years or if each country had it's own good and bad years.

***DATASET*:-**

The dataset we have chosen is “World GDP dataset” from “worldbank.org” . The dataset consists of 199 rows and 21 columns. The rows contain the yearly GDP values of various countires around the world. The columns consist of the country name, country code, indicator name(that is the common indicator used for measuring the GDP), and GDP values from the year 2000 to 2017. The link to the dataset is given below:

[***https://data.worldbank.org/indicator/ny.gdp.mktp.cd***](https://data.worldbank.org/indicator/ny.gdp.mktp.cd)

***INTRODUCTION*:-**

Gross Domestic Product (GDP) is a monetary measure of the market value of all the final goods and services produced in a period of time, often annually or quarterly. Nominal GDP estimates are commonly used to determine the economic performance of a whole country or region, and to make international comparisons.

Economic production and growth, which GDP represents, have a large impact on nearly everyone within that economy. For example, when the economy is healthy, you will typically see low unemployment and wage increases as businesses demand labor to meet the growing economy. A significant change in GDP, whether up or down, usually has a significant effect on the stock market. It's not hard to understand why; a bad economy usually means lower earnings for companies, which translates into lower stock prices. Investors often pay attention to both positive and negative GDP growth when assessing an investment idea or coming up with an investment strategy.

***PRE*-*PROCESSING*:-**

The dataset taken from the world bank website intially contained 62 columns with GDP values of countires starting from the year 1960 and 264 rows with all the countries and also different continent wise totals. During the pre-processing and data cleaning process we had to drop many countries due to the fact that more than 50% of the fields had no data and any analysis carried on such countries would not yield any significant results. Also, we were only able to consider the period 2000-2017 for our project as the preceeding years had not much information and most of the fields were empy.

The very few remaining empty fields after dropping the unimportant rows and columns were filled with succeeding GDP values of each country. The reason for doing this against the conventional method of filling it with the mean value is because of the type of analysis we plan to carry out. Since we analyse any rise or fall of GDP values yearly, filling up the empty fields with a mean value would give rise to unwanted variations in GDP which would in turn give us wron results.

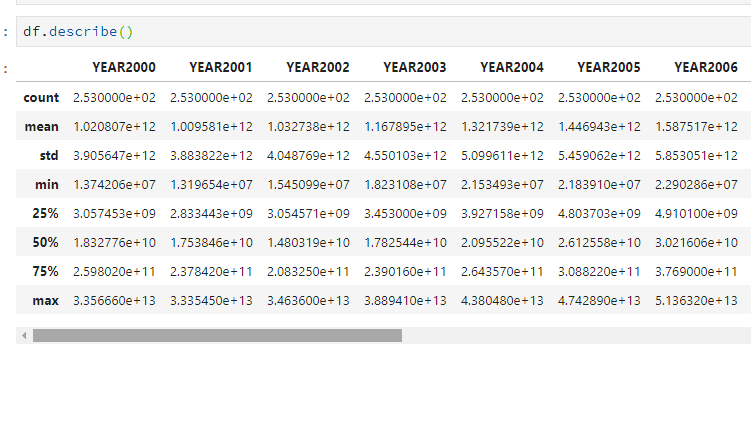
***DESCRIPTIVE ANALYSIS:-***

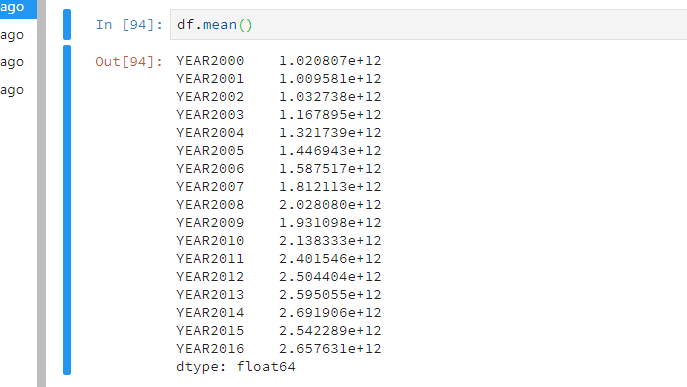
We initially start out with simple and straightforward analysis of the cleaned dataset. We measure the various parameters such as mean, median, maximum value, minimum value, standard deviation, 25th,50th,75th percentiles year-wise. A few of the measured values have been listed in the table below.It gives us an overview of how the GDP’s have changed globally on an average between 2000 and 2016.

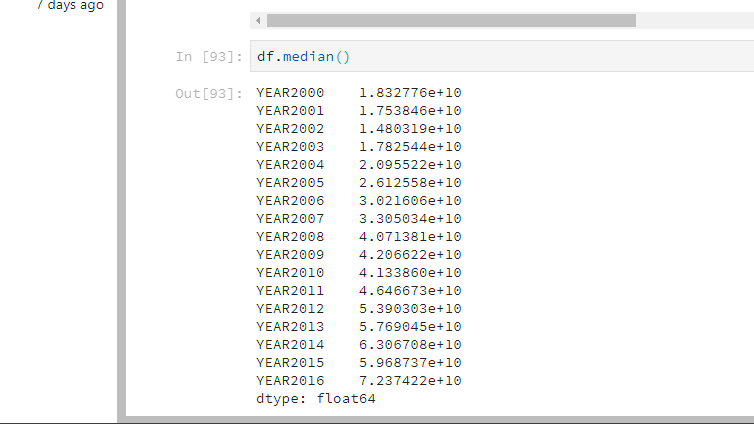
|  |  |  |
| --- | --- | --- |
| **MEASURED VALUE** | **YEAR** **2000** | **YEAR** **2016** |
| Mean | 1.020807e+12 | 2.657631e+12 |
| Standard Deviation | 3.905647e+12 | 8.309881e+12 |
| Minimum | 1.374206e+07 | 3.421888e+07 |
| Maximum | 3.356660e+13 | 7.584510e+13 |
| 25% | 3.057453e+09 | 1.101486e+10 |
| 50% | 1.832776e+10 | 7.237422e+10 |
| 75% | 2.598020e+11 | 1.312160e+12 |

The table describes the various parameters that have been calculated from the dataset.

df.describe() -> this is used to find all the various parameters like mean, standard deviation , min ,max etc. It finds all statistics for the entire dataset (i.e for all the columns present in the dataset)

df.mean()/df.median() -> gives us the mean and median GDP values of those years respectively.





After this, we start analysing the GDP variation for the entire period (2000-2017). We find out the top countries with increasing GDP values (both volume and percentage. The below table shows the top 5 countries according to increase in GDP values.

|  |  |  |
| --- | --- | --- |
| Rank | Country | GDP Increase (USD) |
| 1 | China | 11026353609770 |
| 2 | United States | 9105825000000 |
| 3 | India | 2135344363560 |
| 4 | Germany | 1727485195743 |
| 5 | Brazil | 1400084856748 |

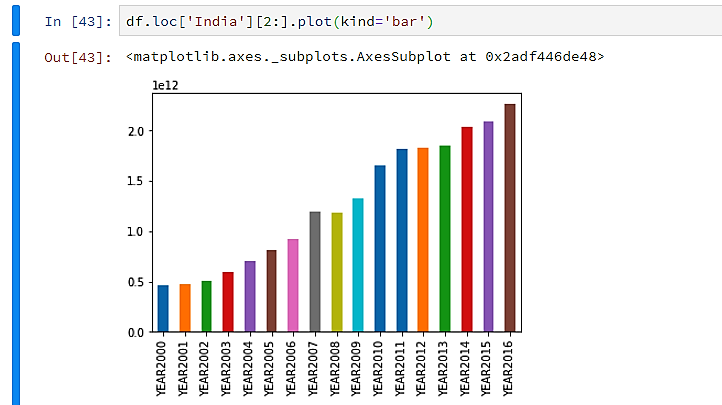
This table provides us information about the top 5 countries according to increase in percentage of GDP.

|  |  |  |
| --- | --- | --- |
| Rank | Country | GDP Increase (%) |
| 1 | Turkmenistan | 13.58 |
| 2 | Angola | 12.6 |
| 3 | Equatorial Guinea | 10.93 |
| 4 | Mongolia | 9.1 |
| 5 | China | 9.1 |

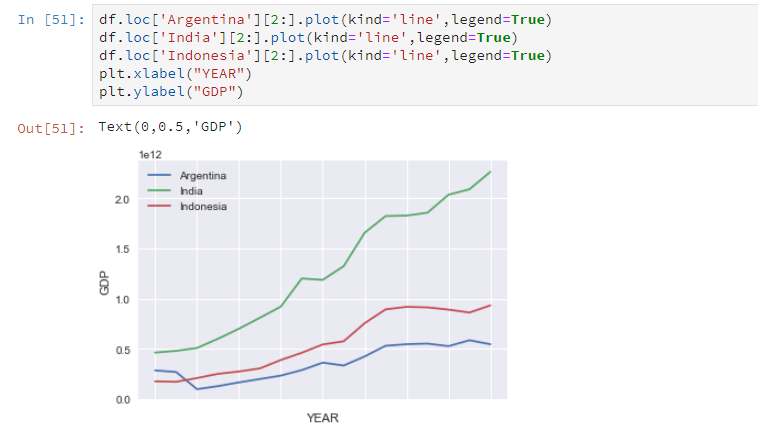
Although most of the countries we analysed had an increasing GDP values when compared between 2000 and 2016, we were able to find two contries which had an overall decrease. The below table shows the country name and percentage of decrease.

|  |  |  |
| --- | --- | --- |
| Country | GDP decrease | |
| Value(USD) | Percentage(%) |
| Japan | 15382715237 | 0.3 |
| Northern Mariana Islands | 42000000 | 3.27 |

We also plot a few sample graphs using the dataset :

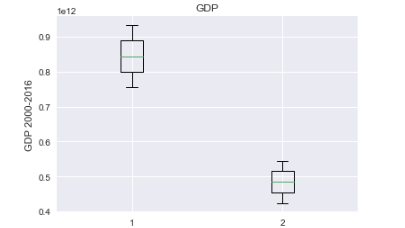


The above graph represents the GDP the country ‘India’ for the various years 2000 to 2016.It is a bar graph representing trends in GDP of a country for multiple years For the above graph we can also add a legend denoting what every block of bar graph represents.

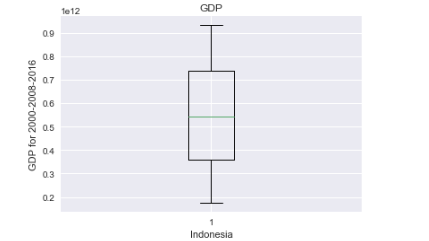


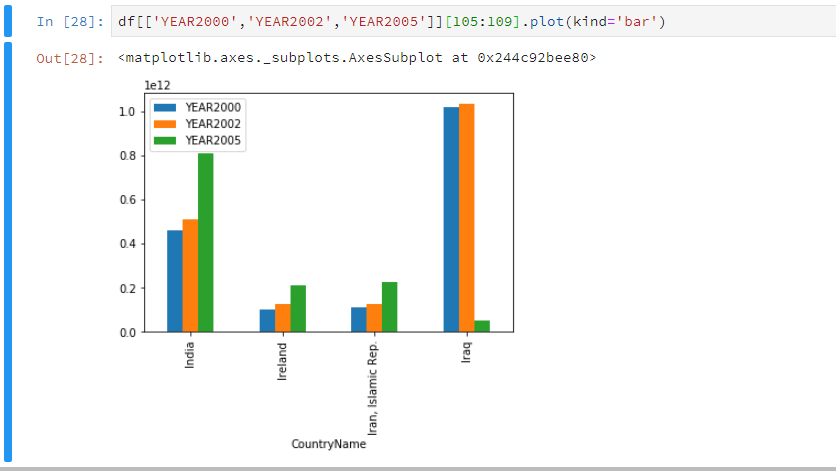
The above graph represents the comparison of GDP between 3 country’s (Ex-Argentina, India, Indonesia) for the timeline 2000-2016 using a line graph. Line graph is best suited to analyse the variations of a particular quantity over a period, (time is taken as one of the measures in most of the cases)

We have also plotted a boxplot for the years 2000-2016 for 2 countries Argentina and Indonesia. When we compare the 2 country’s distribution we can infer about their type of distribution , that is whether they are symmetrically distributed or skewed either on the right or the left.



The below image gives the boxplot for Indonesia for the years 2000,2008,2016 . We can infer from the boxplot that the GDP values of the country Indonesia for the respective years are symmetrically distributed looking at the boxplot diagram.





The above graph gives us information on the GDP of various country’s (Ex: India, Ireland, Iran, Iraq) for different years. This gives us the trends in GDP for a period of time, it helps in comparing country’s over a range of years. Simultaneously we can understand the trends of GDP in various countries around the globe.

After analysing the overall variation of GDP over the given time period, we move on the analysis of the yearly GDP changes. Although most of the country’s GDP increases over the time period, each country has its own rises and falls. We try to analyse this with the data in hand. The below table lists the good and bad years for a few important countries.

|  |  |  |
| --- | --- | --- |
| Country | Years | |
| India | Good | 2000-2006, 2008-20017 |
| Bad | 2007 |
| Pakistan | Good | 2002-2007, 2009-2017 |
| Bad | 2000, 2001, 2008 |
| United States | Good | 2000-2007, 2009-2017 |
| Bad | 2008 |
| China | Good | 2000-2017 |
| Bad | None |

Also, we find out a year-wise count of the number of countries whoose GDP increased in that year. The following table provides the top 5 good and bad years and their respective count.

|  |  |
| --- | --- |
| YEAR | Number of countries whoose GDP increased |
| 2006 | 195 |
| 2003 | 193 |
| 2005 | 193 |
| 2004 | 191 |
| 2010 | 190 |

|  |  |
| --- | --- |
| YEAR | Number of countries whoose GDP increased |
| 2014 | 55 |
| 2008 | 57 |
| 2000 | 118 |
| 2011 | 133 |
| 2015 | 134 |

As we can see from the above data, the years 2008 and 2014 have been bad years for a major portion of countries around the world. A possible reason for this could be the fall in value of the US dollar.

Finally, we found out a bunch of countries which never had a bad year, i.e., their GDP kept increasing through out the analysed time period. They are **China, Bangladesh, Jordan, Kenya, Cambodia, Lebanon, Panama and Vietnam.**

***TYPE OF DISTRIBUTION:-***

When we plot the data of respective country’s against the years starting from 2000 to 2018, we see that is is **negatively skewed,** i.e., the graph increases from left to right. Since the graph is negatively skewed, we can draw a conclusion that most of the countries around the world are increasing their productivity which in turn is increasing their GDP levels.

***CONFIDENCE INTERVAL:-***

We have built the confidence interval for a particular year 2015 , with significance level 1% and sample size of 45.We have built the confidence interval by using the z-score value and the margin of error.

By using the respective formulae we calculate the upper bound the lower bound for the year 2015.

Confidence interval:

We construct a 90% confidence interval for mean of a sample set taken from the year 2015.

Sample mean: 2.83x1012

Sample standard deviation: 8.45x1012

Results:

Lower bound of mean: 7.59x1011

Upper bound of mean: 4.90x1012

***HYPOTHESIS TESTING:-***

We have chosen the to conduct our hypothesis testing for 95% rejection region (significance level = 5%).

Problem statement :- *“mean GDP of countries in the year 2016 is more than twice of the mean GDP of countries in the year 2006”.*

In order to perform the hypothesis testing we consider our Null Hypothesis (Ho) as mean GDP of countries in 2016 is not more than twice the mean GDP of countries in 2006 and Alternate Hypothesis (Ha) as mean GDP in 2016 is more than twice the mean GDP in 2006.

We are carrying out a single-tailed (right tailed) hypothesis testing.

Based on the result of the hypothesis testing we have concluded that the Null Hypothesis (Ho) can’t be rejected and thus we have said that Ho is plausible.

Hypothesis testing.

Following hypothesis is to test if the mean gdp of countries in the year 2016 is more than twice the mean gdp of the countries in the year 2006.

MUx denotes the sample mean gdp of countries in year 2006.

MUy denotes the sample mean gdp of countries in year 2016.

Sample size = 45

Null hypothesis : 2\*MUx - MUy <= 0

Alternate hypothesis : 2\*MUx - MUy > 0

actual z value : 1.64

hypothesis z value : 0.54

The result of the hypothesis test is:

Null hypothesis is probable.

***CONCLUSION* :-**

The gross domestic product (GDP) is one of the primary indicators used to gauge the health of a country's economy. It gives a bird-eye view how the economy of the country is performing.

If the GDP increases from year to year it means that the country is doing well with respect to the final productions of various departments and sectors and consumptions by the citizens of the country.

We can also conclude that if the GDP of country is doing pretty good then the citizens of the country are all also benefitting in terms of job opportunity and the incomes being offered to the employees.

If the GDP growth is good for a country then the country is progressing towards its development in all fields like medical, technology, sports, travel and tourism etc.