**STATISTICAL ANALYSIS USING SAS ENTERPRISE GUIDE AND DASHBOARD DEVELOPMENT USING TABLEAU**

By

TAIWO OLORUNTELE

(*Student id* @00609735)

# Principle of Data Science Module

School of Computing, Science & Engineering

University of Salford, Manchester.

January 13, 2022

# 1. INTRODUCTION:

In this project, I shall examine the economic health of various countries and how this impacts their citizens using data from the World Development indicators in the World Bank database (Worldbank.org 2015) for a period of 15 years from 2006 to 2020 which I believe is a good metric since the world has changed significantly during from the last decade.

For the selection of my countries, I have gone with picking based on GDP as my main target as I believe is a great metric for determining how well a country is doing. On this basis, I looked at Statista and found 5 of the best performing European countries, 4 of the worst African countries, and 4 of the best African countries, which happen to be from the MENA (Middle East and North Africa) region.

These countries are listed below as well as a chart showing the GDP metrics over 15 years.

1. Germany
2. France
3. Spain
4. Netherlands
5. Italy
6. Algeria
7. Tunisia
8. Morocco
9. Egypt

10. Burundi
11. Gambia
12. Lesotho
13. Guinea-Bissau

## GDP over years

**Year**
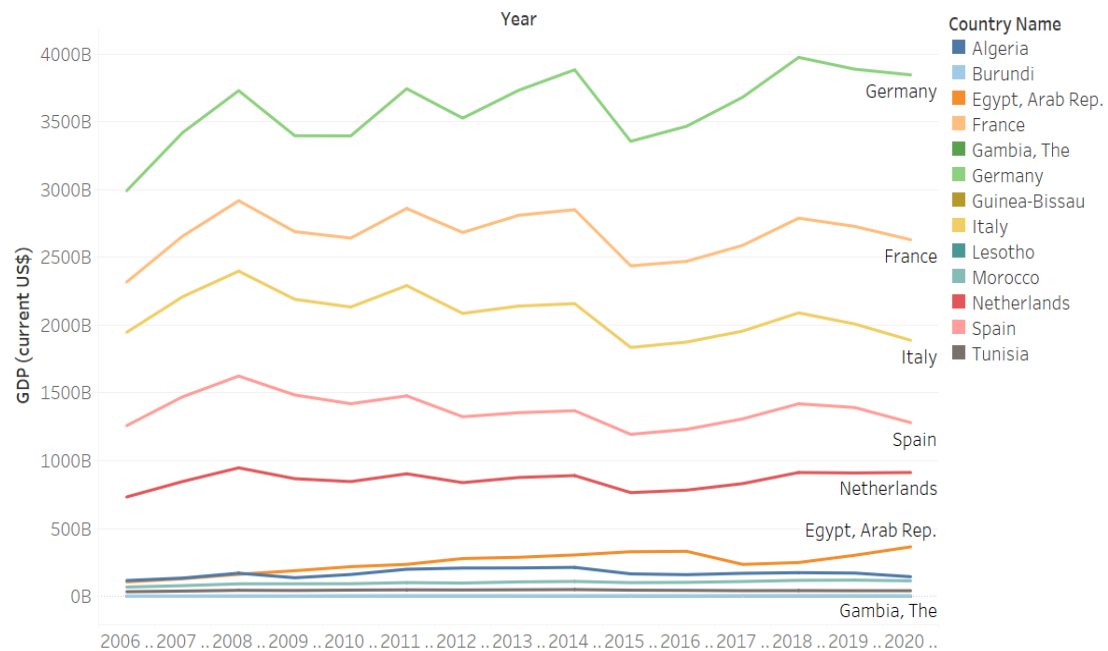


The data reveals that Germany leads, as well as the European countries, with Egypt being the best performer in Africa and Gambia being the worst.

## 2. BACKGROUND RESEARCH:

What is a dashboard? Essentially, a dashboard is a tool that tracks, analyzes, and displays KPIs, metrics, and critical data points. The goal of displaying these data points is to make the information more easily understandable to the average user. Dashboards are a great business intelligence tool that can be used to make well informed decisions (Computing 2020).  In this manner, various trends over a particular time period or various attributes can be analyzed at a glance, for instance data over a specified period of time, top performing ads on a site, monthly visitors on a site, etc. As I have shown above, I can also easily calculate the GDP of my selected countries.

There are various software used for dashboard creation including Tableau, Microsoft Power BI, Microsoft Excel but for this project, I have utilized Tableau which I found to be the best and most interactive to use.

What makes a good dashboard? You must create dashboards that are easily readable and easy to use. If your audience has trouble making sense of them, the entire purpose is defeated. To avoid confusing the audience, you must be explicit with the outcome you are aiming for. You also need to pay attention to the colors because you don't want to blind your audience. Colors play an enormous role in your presentation, picking the right blend enables your audience with easy identification.

## Dashboard methodologies:

It is also critical to plan your dashboard well in advance. You need to be familiar with the different types of charts and pick the one that best suits the story you're telling. If you want to show trends, you'll most likely want to use a bar chart or line chart while if you want to show composition, then a pie chart is more appropriate. Noting as well,  use of a bar chart for too much

components of a total value can make your chart look jumbled up and unreadable.

Numbers:- Giving context to your numbers matter a lot. For example, you'd want to highlight bad numbers as Red and good numbers as Green since Red is known to be a colour to depict danger and Green for prosperity.

Grouping:- Having similar KPIs together is key, if you have to work with multiple. You could have sales numbers on one dashboard and region related numbers on another dashboard, it wouldn't make much sense during presentation to switch from dashboard to dashboard, when you could simply have related ones on single dashboard.

Labels:- Use of labels is also important, as you do not want to start to figure out what numbers relate to what after you're done building, when you could easily have done so from the start.

Animations:- An overlooked chart method is the animated bar chart, which could show movement over the years, for example plotting sales number of artistes over 10-20 years, it would be useful to see how they performed overtaking eachother and when it happened.

# 3. EXPLORATION OF DATASET:

As stated earlier, this project is based on economic health situation of various companies and its impact on its citizens. After carefully going through the databases available on the World Bank side, I decided to go with a mix of the Jobs section and Health section as I believe those are two major components of the economy. Jobs focus on contributing to the GDP majorly and Health on how much the government is committed to keeping its citizens in good condition. All variables I picked ranged over 15 years from 2016 to 2020 and are listed below:

1. GDP (current US$)
2. Access to electricity, urban (% of urban population)
3. Access to electricity (% of population)
4. Antiretroviral therapy coverage (% of people living with HIV)
5. Children (0-14) living with HIV
6. Children (ages 0-14) newly infected with HIV
7. Current health expenditure (% of GDP)
8. Incidence of tuberculosis (per 100,000 people)
9. Incidence of malaria (per 1,000 population at risk)
10. International tourism, number of arrivals
11. Newborns protected against tetanus (%)
12. Labor force, female (% of total labor force)
13. Labor force, total
14. Exports of goods and services (% of GDP)
15. Fixed telephone subscriptions
16. Fixed broadband subscriptions
17. Mobile cellular subscriptions
18. Urban population
19. Individuals using the Internet (% of population)

For proper analysis, I have had to Transpose the data in SAS Enterprise Guide, to get the columns in a better format. Also, a few data was missing for some years which I imputed based on the Average of the last 3 years after my preparation and cleaning, I was left with no empty spaces. Here's an excerpt of the data after Transposing, which I saved as an Excel file then imported to Tableau:

| | Country Name | Years | GDP (current US$) | Access to electricity, urban (% | Access to electricity (% of popu | Antiretroviral therapy coverage | Children (0-14) living with HIV | Children (ages 0-14) newly infec | Current health expenditure (% of | Incidence of tuberculosis (per 1 | Incidence of malaria (per 1,000 | Intern tou num |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Algeria | 2006 | 117027000000 | 99 | 99 | 22 | 200 | 100 | 3.35510325 | 79 | 0.000506097 | 1 |
| 2 | Algeria | 2007 | 134977000000 | 99 | 99 | 23 | 200 | 100 | 3.82141781 | 78 | 0.012955969 | 1 |
| 3 | Algeria | 2008 | 171001000000 | 100 | 99.3 | 23 | 200 | 100 | 4.20188236 | 74 | 0.001470659 | 1 |
| 4 | Algeria | 2009 | 137211000000 | 99.50669098 | 98.81757355 | 29 | 200 | 100 | 5.35939837 | 77 | | 1 |
| 5 | Algeria | 2010 | 161207000000 | 99.55960083 | 98.82183838 | 38 | 200 | 100 | 5.11716986 | 78 | 0.000473231 | 2 |
| 6 | Algeria | 2011 | 200013000000 | 99.61138916 | 98.84799194 | 41 | 500 | 100 | 5.26748085 | 73 | 0.000464402 | 2 |
| 7 | Algeria | 2012 | 209059000000 | 99.25673759 | 98.76466047 | 47 | 500 | 100 | 6.00050974 | 73 | 0.025048469 | 2 |
| 8 | Algeria | 2013 | 209755000000 | 99.70097351 | 98.97725677 | 56 | 500 | 100 | 6.03576279 | 68 | 0.003571173 | 2 |
| 9 | Algeria | 2014 | 213810000000 | 99.73947906 | 99.07403564 | 64 | 500 | 100 | 6.54721403 | 72 | 0 | 2 |
| 10 | Algeria | 2015 | 165979000000 | 99.77484894 | 99.18688965 | 71 | 500 | 200 | 6.97849178 | 75 | 0 | 1 |
| 11 | Algeria | 2016 | 160034000000 | 99.8078537 | 99.40287018 | 77 | 500 | 200 | 6.60778189 | 70 | 0 | 2 |
| 12 | Algeria | 2017 | 170097000000 | 99.83929443 | 99.61513519 | 79 | 580 | 200 | 6.38032866 | 70 | 0 | 2 |
| 13 | Algeria | 2018 | 174911000000 | 99.8981781 | 99.641922 | 83 | 660 | 200 | 6.2184267 | 69 | 0 | 2 |
| 14 | Algeria | 2019 | 171767000000 | 99.8 | 99.5 | 85 | 750 | 200 | 6.2 | 61 | 0 | 2 |
| 15 | Algeria | 2020 | 145009000000 | 99 | 99 | 85 | 850 | 200 | 6.346 | 59 | 0 | 2 |
| 16 | Burundi | 2006 | 1273375020 | 51.59951782 | 2.66 | 7 | 26000 | 2900 | 9.94232273 | 181 | 235.6769652 | |
| 17 | Burundi | 2007 | 1356199365 | 51.80656433 | 4.775167942 | 11 | 25000 | 2800 | 10.5808115 | 169 | 222.9154949 | |
| 18 | Burundi | 2008 | 1611835902 | 39.6 | 4.8 | 14 | 23000 | 2500 | 9.23542786 | 160 | 222.7746531 | |
| 19 | Burundi | 2009 | 1781455092 | 52.25225067 | 5.40973711 | 18 | 22000 | 2400 | 10.35023212 | 152 | 220.6495356 | |
| 20 | Burundi | 2010 | 2032135247 | 49 | 5.3 | 24 | 20000 | 1900 | 11.27638912 | 144 | 210.1978813 | |
| 21 | Burundi | 2011 | 2235820868 | 52.71091843 | 6.106936455 | 28 | 19000 | 1800 | 10.06996346 | 138 | 184.1450365 | |
| 22 | Burundi | 2012 | 2333308099 | 58.5 | 6.5 | 32 | 18000 | 1700 | 8.52668476 | 133 | 153.9276694 | |
| 23 | Burundi | 2013 | 2451625333 | 52.4 | 6.9 | 36 | 16000 | 1500 | 8.14256191 | 128 | 140.5884216 | |
| 24 | Burundi | 2014 | 2705783272 | 52.1 | 7 | 42 | 15000 | 1300 | 6.88927555 | 126 | 141.5075585 | |
| 25 | Burundi | 2015 | 3104394858 | 57.14767456 | 8.403089523 | 48 | 14000 | 1200 | 6.57857561 | 122 | 165.5009225 | |
| 26 | Burundi | 2016 | 2732808557 | 58.69766998 | 9.251799583 | 60 | 13000 | 930 | 7.49265718 | 118 | 225.743378 | |
| 27 | Burundi | 2017 | 2748180474 | 61.8 | 9.3 | 71 | 12000 | 960 | 7.51763725 | 114 | 250.2725129 | |
| 28 | Burundi | 2018 | 2668495743 | 61.54690933 | 10.59861469 | 78 | 11000 | 980 | 7.74165535 | 111 | 250.2724964 | |
| 29 | Burundi | 2019 | 2631434363 | 62.74242401 | 11.0647974 | 86 | 10000 | 920 | 7.8 | 107 | 241.677 | |
| 30 | Burundi | 2020 | 2841786382 | 65.433 | 13.654 | 88 | 9700 | 950 | 7.9333 | 103 | 247 | |

# 4. INVESTIGATION OF DATA WORKFLOWS & PROPOSAL FOR DESIGN OF DASHBOARD

Dashboard aim:

1. To find relationship between Export of Goods and Services and GDP
2. To find relationship between International Tourism and GDP
3. To find the gender gap occurrence between countries
4. To find how much of the GDP is spent towards Health matters
5. To find the relationship between Antiretroviral Therapy offered to children and their HIV rates

**Selection of Charts:** for most of this research I have had to use bar charts and line graphs, as most of the data operates on time series spread over years. I made use of the dual axis charts for my comparisons what I wanted to see the relationship between GDP and current health expenditure, GDP and international tourism, GDP and exports of goods and services, antiretroviral therapy and the prevalence of HIV in children ages 0 to 14. I have also made use of the map to show the positions of my selected countries and get some data off the specified KPI.

**KPIs used:** as specified earlier above the variables I picked she used as my kpi I will further go on with the explanation of my most important ones.

**Dashboard structure:** For my dashboards, I have decided to create two(one representing Health and the other based on economy) as I have examined my data and as said earlier, it wouldn't make too much sense to have a cluttered dashboard, which is why I decided to go with two.  For both, I have decided to apply a Country Name filter to easily scan through trends of the countries year by year. By doing this, a country is selected one by one and the values change along with it to represent.

1. GDP:  Known as gross domestic product is one of the major indicators to determine how well a country is doing economically. When this is growing, workers and businesses are generally better off than when it is

slowing down. It's constantly be defined how's the total value created in an economy over a particular time.

## GDP over years
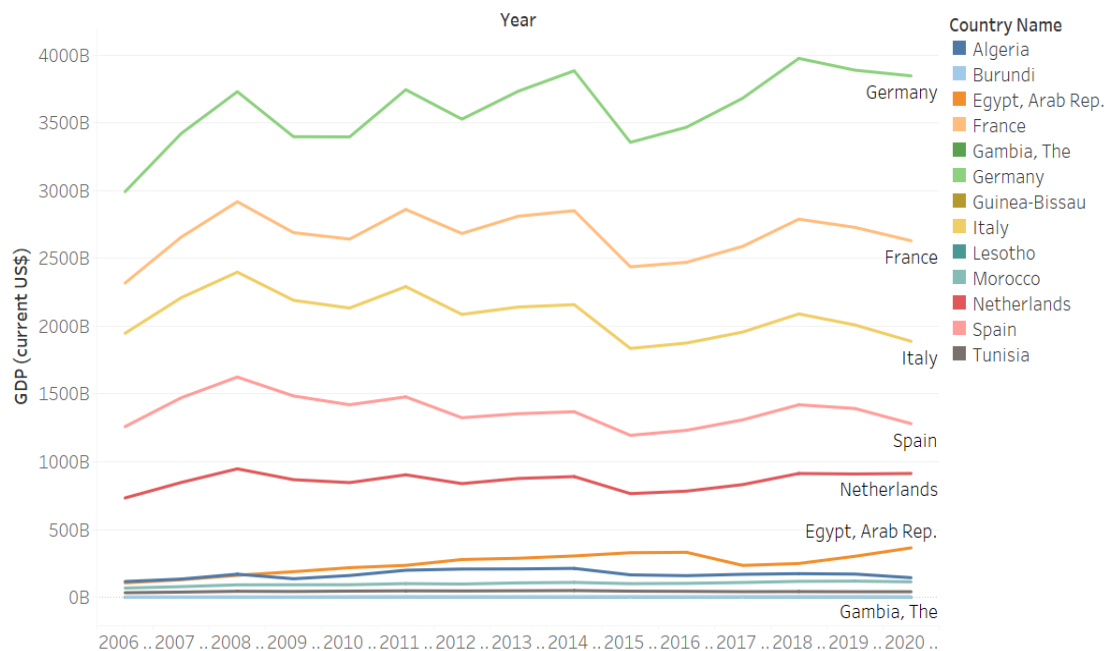


2. **Access to electricity, urban (% of urban population):** This specifies the percentage of the urban population that has access to electricity. Access to electricity has proven to be a great indicator of how well a country has progressed because this enables its people to healthier living, internet and televisions which is key to getting more information and has helped boost businesses. We have seen more startups and businesses go online. Only with constant supply of electricity they can thrive. Globally, the percentage of people with access to electricity has been steadily increasing over the last few decades. In 1990, around 71% of the world's population had access and increased to 87% in 2016. (Data 2020)

## Electricity



© 2022 Mapbox © OpenStreetMap
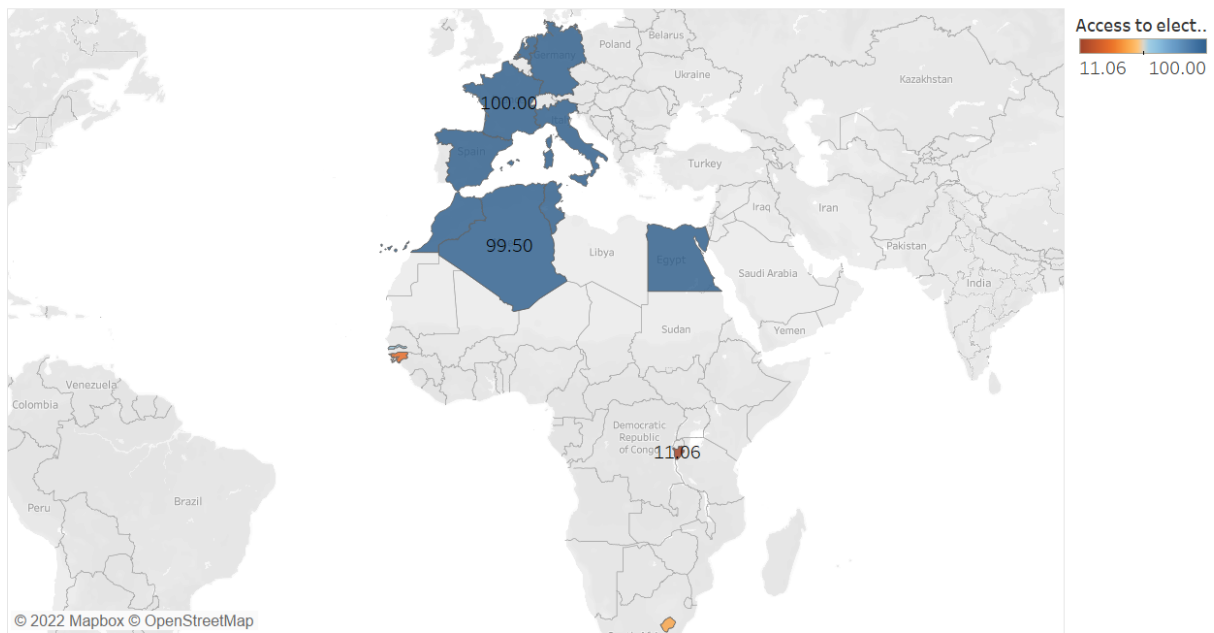
3.  Antiretroviral therapy coverage (% of people living with HIV): this specifies the percentage of the total amount of people Living with HIV that's how access to antiretroviral therapy. I believe the government has a lot to do with providing as this is not something that can easily be gotten. The more people who have access to this , the more they can save the lives of the infected As HIV  still seems to be a major cause of death. On the global front, between 55.9 million–110 million people have become infected with HIV since inception and between 27.2 million–47.8 million of this number have died from AIDS-related illnesses. In 2020, 73% of all people living with HIV were accessing treatment. 74%  of adults aged 15 years and older living with HIV had access to treatment, as well as 54% of children aged 0–14 years with a suppression rate of 90% (Unaids.org 2021)

## ART and Child HIV
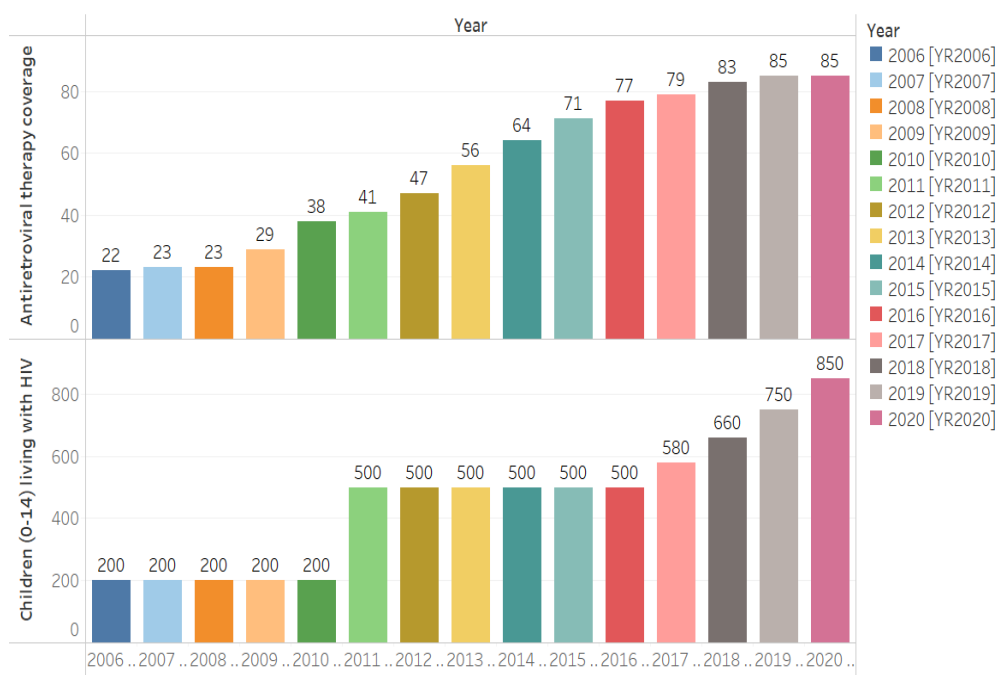


4. **Children (0-14) living with HIV:** this is the number of children who happens to be infected with HIV within the age range of 0-14. This number encompasses the children who were born with the virus and those who happens to have contacted it's one way or the other.

5. **Children (ages 0-14) newly infected with HIV:** This metric shows how fast or how slow children are being infected and would help with identifying what attention needs to be transferred to. This has to be kept at a very low rate as children are more or less the  backbone of the future economy. The UN Joint Program on HIV/AIDS (UNAIDS) estimates that about 85% of 2.6 million children born with the infection are from sub-Saharan Africa e.g Burundi, Lesotho, Guinea etc . About 650,000 children are living with HIV/AIDS and approximately 1000 infected infants are born every day in Africa, which is a staggering rate and has to be looked into by the government. (Marum 2022)

## Newly HIV Infected Children
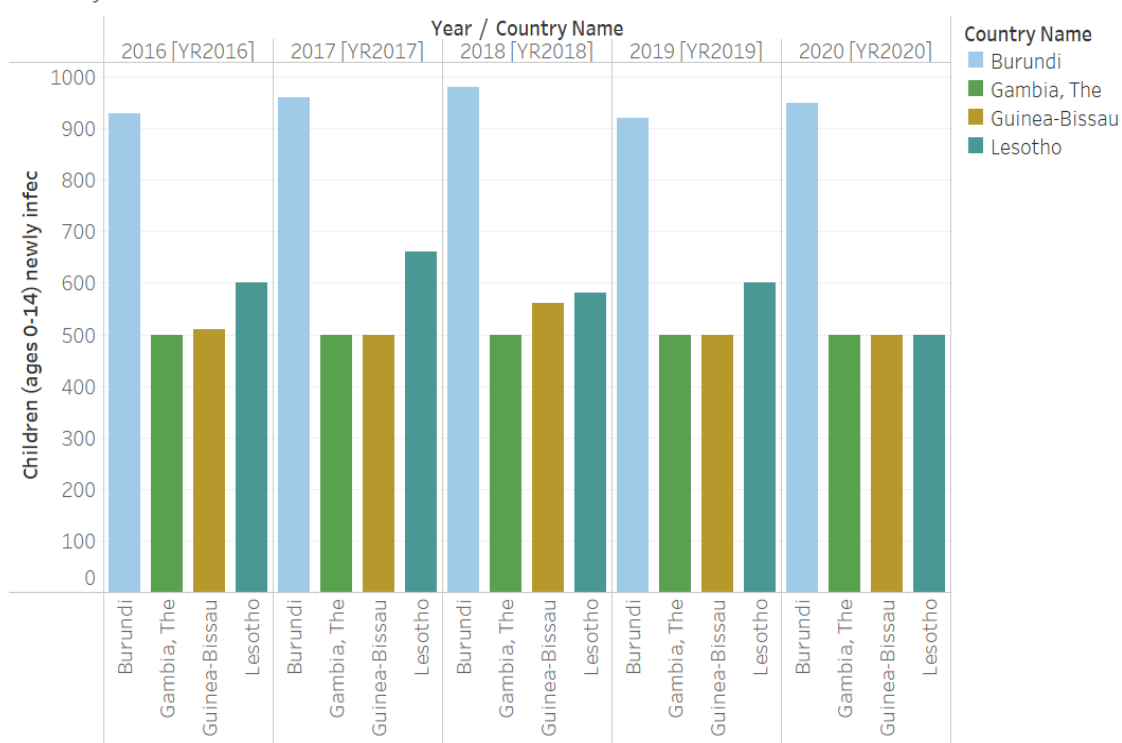


6. Current health expenditure (% of GDP):  this shows how much of the GDP the Government is dedicating through the healthcare system. As I believe with more people healthy they will be able to work and give their best to the economy. An analysis was carried out with data collected from the Bureau of Economic Analysis and the Bureau of Labor Statistics for the years 2003–2014 and suggested a positive correlation between health expenditure, GDP and labor productivity as expected. An increase in healthcare expenditure has a positive relationship with economic performance. (Raghupathi 2020)

## GDP and Health Expenditures
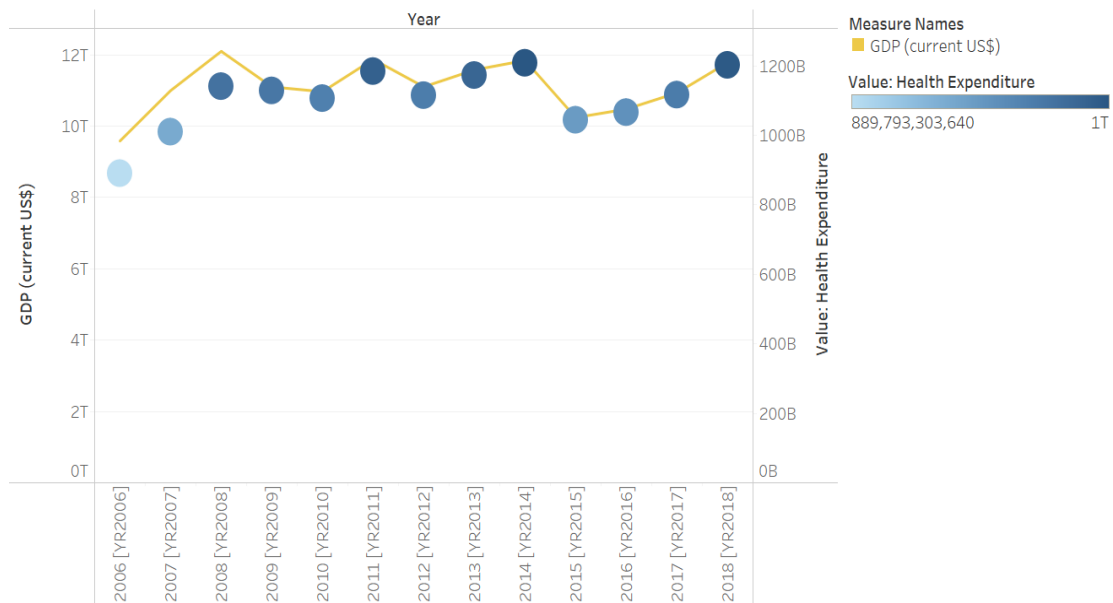


7. **International tourism, number of arrivals:** Tourism is perceived to be a great indicator of foreign exchange, certain countries have leveraged the influx of foreigners into their countries and used it as a measure to generate more income for the government. It does make sense so use it as an extra source and also promotes exposure about the country. Taking a look at 2016, Tourism contributed to GDP growth of 3.1% and created 6 million additional jobs in the sector, totalling 292 million which is equal to 1 in 10 jobs in the economy. It also generated US$7.6 trillion (10.2% of global GDP. (Destinationsinternational.org 2018)

## Tourism numbers for 2016

**Country Name**

International tourism, number of

| | |
|---|---|
| Algeria | 2,039,000 |
| Burundi | 187,000 |
| Egypt, Arab Rep. | 5,399,000 |
| France | 203,042,000 |
| Gambia, The | 450,000 |
| Germany | 35,555,000 |
| Guinea-Bissau | 45,200 |
| Italy | 84,925,000 |
| Lesotho | 1,196,000 |
| Morocco | 10,677,000 |
| Netherlands | 15,828,000 |
| Spain | 115,561,000 |
| Tunisia | 5,724,000 |

**Country Name**
- Algeria
- Burundi
- Egypt, Arab Rep.
- France
- Gambia, The
- Germany
- Guinea-Bissau
- Italy
- Lesotho
- Morocco
- Netherlands
- Spain
- Tunisia

8. **Labor force, female (% of total labor force):** This focuses on gender gap, it is believed it is very wide in most countries but as varies due to different factors. Also, there have been debates about how hiring more females professionally has helped GDP significantly

## Female labor over years
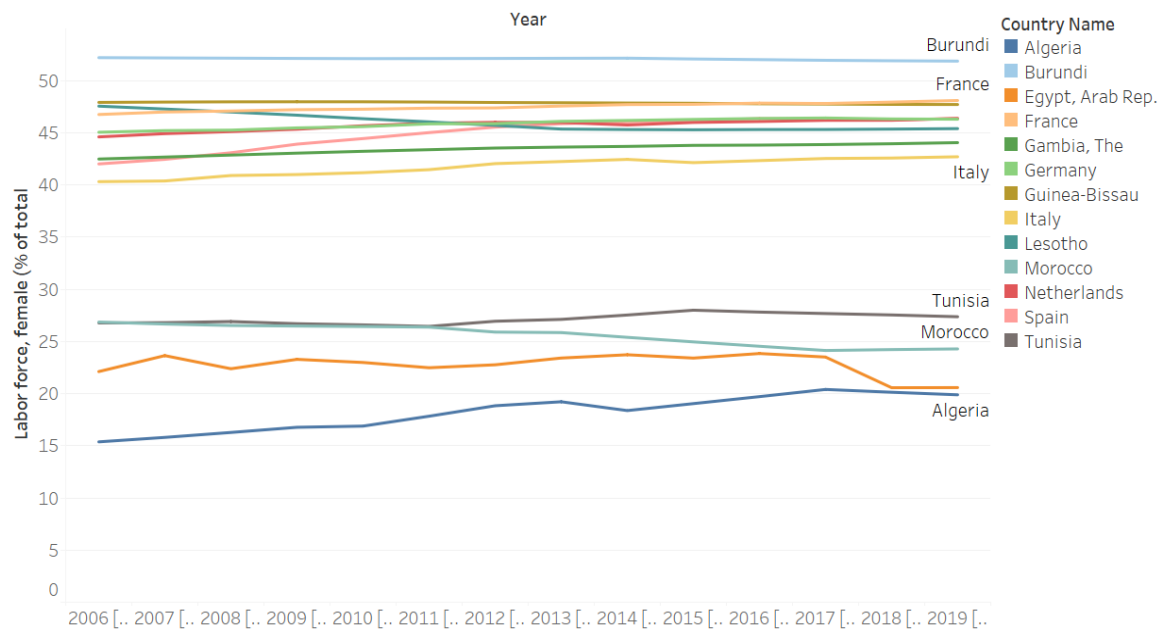


9. **Exports of goods and services (% of GDP):** This is one of the major components of a country's GDP. Most countries rely on this as not one single country has everything they need. For example, Nigeria exports a lot of crude oil to other countries and  this account for a great percentage of their GDP. Exports typically facilitate international trade and stimulate domestic economic activity by creating employment, production, and revenues. When there are more exports, it means that there is a high level of output from a country's factories and industrial facilities, as well as a greater number of people that are being employed in order to keep these factories in operation

## Export comparison with GDP
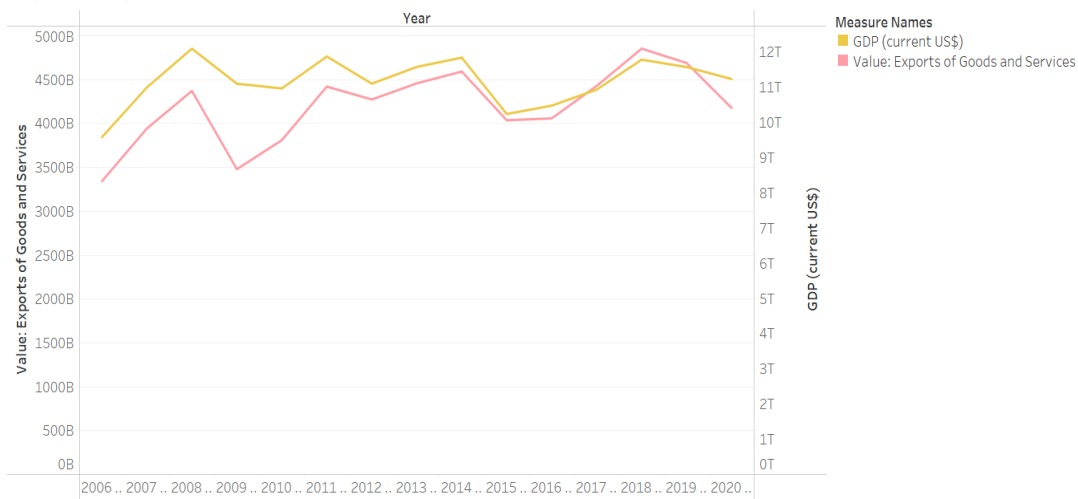


10. Urban population: This accounts for the number of people that do not live in rural settlements and this is a great indicator for growth as we have more people living in urban settlements in developed countries which their GDP is higher than those of poor countries with more people living in rural settlements. As at 2020, averagely 55% of the world's population live in urban settlements. It is estimated that by 2050, the number of urban population would've doubled to about 7 of 10 people living in cities. It is seen that 80% of global GDP is generated in cities and some governments have realised this, thereby ramping up urbanization to in turn continue to contribute to sustainable growth if managed well by increasing productivity, allowing innovation and new ideas to emerge.

# 4.DISCUSSION:

Going through my dashboards and looking at both Tourism and Exports from Gambia, I was able to take note of an uptick in International Tourism from 2015 which drove the GDP higher significantly and how the export value also fell drastically from 2020. On to do more research on Tourism, I discovered the government has had a role to play as they have been most interested in the construction of hotels in the 4- and 5- star class ranges. The River Gambia is also open for the development of cruising, sport fishing, and eco-tourism resorts. International flight options are growing, with Air Senegal, TAP and Turkish Airlines expanding their routes to include The Gambia in the past two years, joining Brussels Airlines and Royal Air Maroc as the major international operators in The Gambia. I also went on to gather more information about why its exports fell so bad and I could relate it to Covid outbreak in which there was not much access to the outside world, of which its Export value wasn't so strong beforehand.

## GDP and Tourism
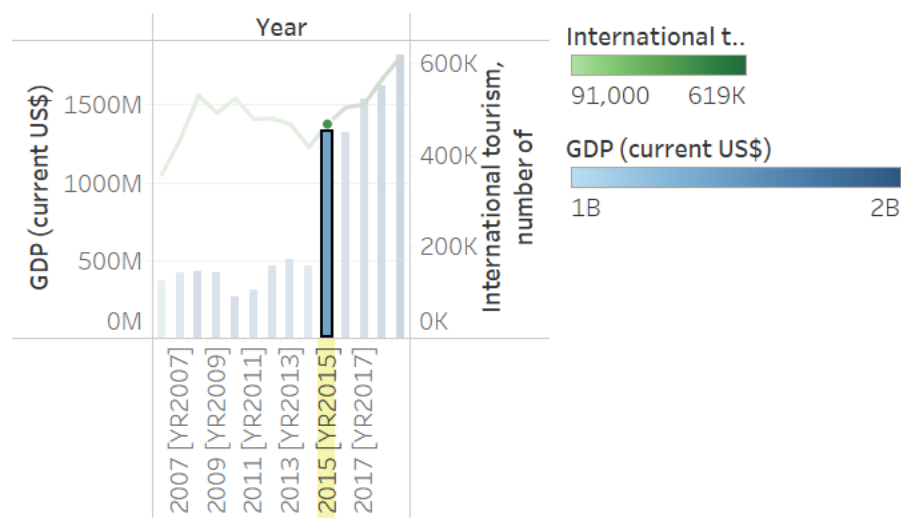
## Export comparison with GDP
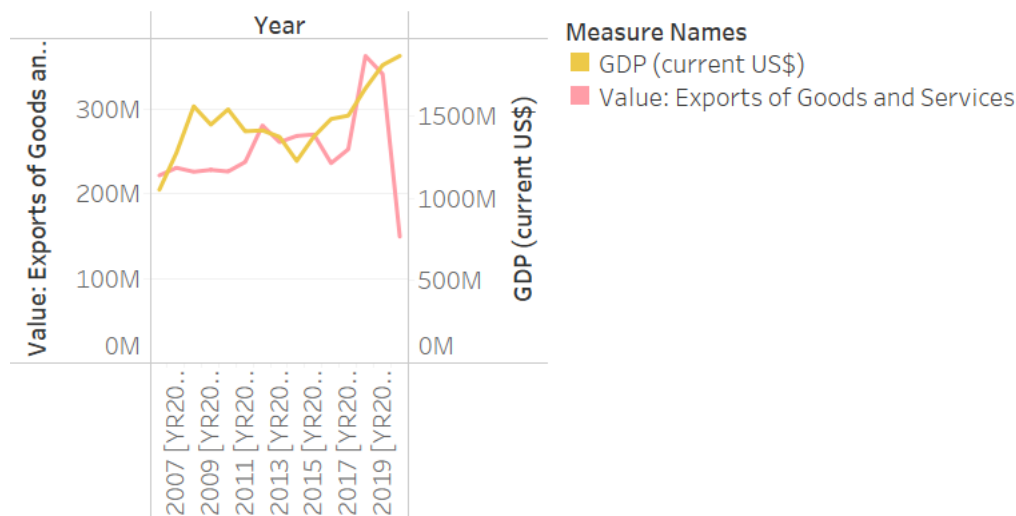


Going through the other dashboard, I also noticed that the amount of access to Anti Retroviral Therapy kept increasing year by year as the number of children who are HIV positive kept increasing, which shows the government's awareness to the situation and will to battle it but for Burundi and Lesotho, the Anti Retroviral Therapy and number of infected children was inversely proportional. The ART coverage was steady increasing while the number kept reducing, which shows that despite the rates reducing, the government didn't relent on efforts to increase the ART coverage. Once again, I went on to do some research on this and found out that Burundi launched a National HIV/AIDs Strategy (NHAS). The project's main aim in relation to this was "increase the utilization of a selected set of preventive services, among groups highly vulnerable to, or affected by HIV/AIDS." Commercial sex workers, Burundians aged 15-49 with multiple sex partners in the past year, and pregnant women with HIV were all targeted by the project.

**Dashboard:**

With this dashboard, I am able to inspect four aspects at the same time. The female labor force over time, the amount of individuals using the internet in

2006, the relationship between GDP and Tourism over the years and Export comparison with GDP over the years.

I set this up in such a way that changing the country name on the top right corner would apply to all worksheets asides the Individuals Using the Internet. This way, I would be able to view each country at a go and look at their variations, rather than having to go select each country image one by one as that would take more time and effort.
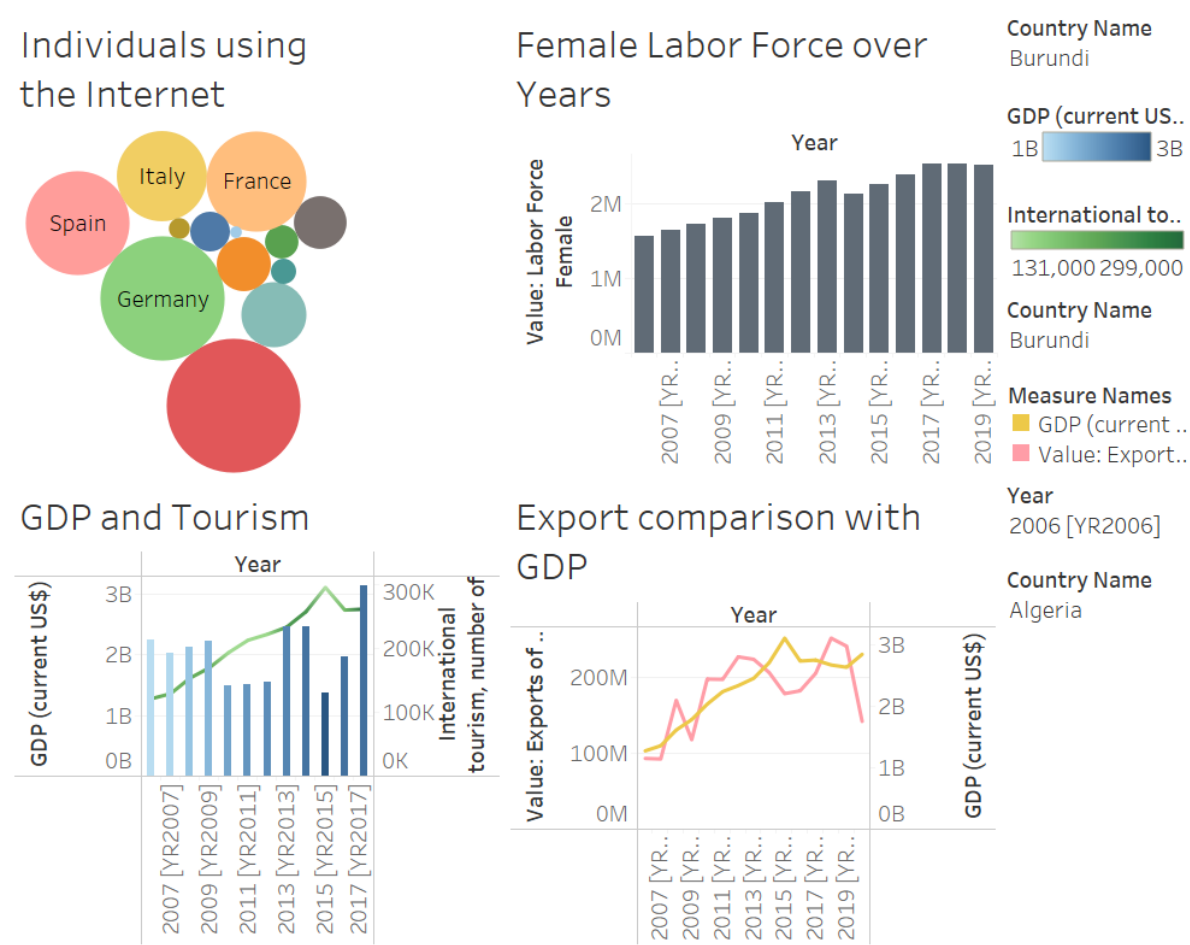


For my forecast, I have decided to go with GDP based on the next 5 years and it is shown below:

## GDP Forecast until 5 years



**Country Name, Forecast indicator**
- Algeria, Actual
- Algeria, Estimate
- Burundi, Actual
- Burundi, Estimate
- Egypt, Arab Rep., Actual
- Egypt, Arab Rep., Estimate
- France, Actual
- France, Estimate
- Gambia, The, Actual
- Gambia, The, Estimate
- Germany, Actual
- Germany, Estimate
- Guinea-Bissau, Actual
- Guinea-Bissau, Estimate
- Italy, Actual
- Italy, Estimate
- Lesotho, Actual
- Lesotho, Estimate
- Morocco, Actual
- Morocco, Estimate
- Netherlands, Actual
- Netherlands, Estimate
- Spain, Actual
- Spain, Estimate
- Tunisia, Actual
- Tunisia, Estimate

# CONCLUSION

At the end of this, I have been able to create various dashboards with which a user can play around with to get different insights on most of my KPIs used. The user can easily switch from the health of the citizens to the GDP impacts. With dual axis, I was able to find out how the number of international tourists have added to its GDP, correlation between the country's exports of goods, its female labor contribution all over 10 years. To get the exact figures for the percentages of female labor and access to electricity, I made use of calculated fields.

Understanding each KPI used was very key and I made sure to do a thorough research on each of them. Doing this, I was able to get extra information for myself that I had no idea about before I took on the project.

# STATISTICAL ANALYSIS

## 1. INTRODUCTION

This task focuses on the same specified data as visualized above, from the World Bank database. Statistics embodies the collection and analysis of large amounts of data to discover certain patterns with the numbers. Statistics is applied in everyday use in retail, health, aviation, education.

For proper Statistical analysis, the following needs to be done. You must understand the nature of the data to be analyzed, to know what analysis that could be done and set your objectives beforehand. As well, you would need to explore the relationship between the various aspects of the data.

## 2. BACKGROUND RESEARCH

In Statistics, there are very many terms and definitions one needs to be familiar with so below, I have explained the major parameters used.

**MEAN**:

 This can easily be defined as the mathematical average of two or more numbers.

**Mean = 6**

2    7  9

0  1  2  3  4  5  6  7  8  9  10  11
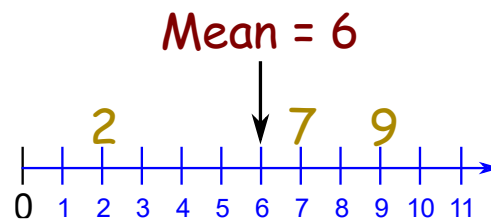
MEDIAN:

This can be defined as the midpoint of a range of variables sorted in a descending or ascending order

10  11  13  15  16  23  26

middle number
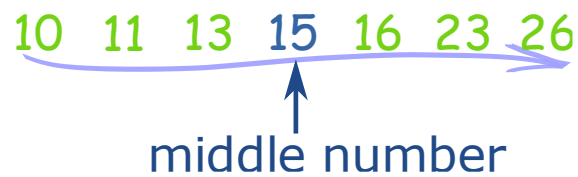
SKEWNESS:

This can be defined as a measure of the symmetry of a distribution with the highest point of its distribution being its mode. It should be noted if its unimodal i.e it has one mode. If so, it can then be skewed towards the right(positive) or towards the left(negative) as represented with images below.
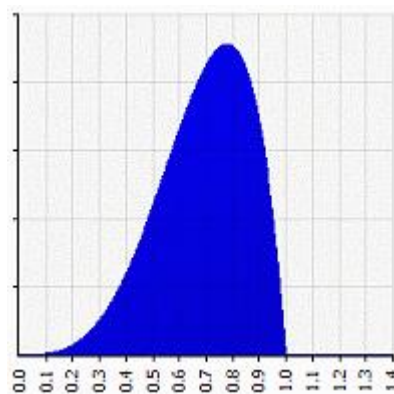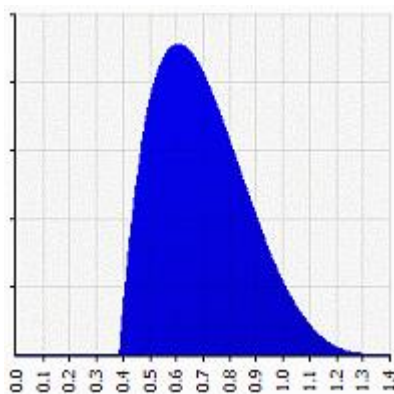
If skewness is less than −1 or greater than +1, the distribution is highly skewed.

If skewness is between −1 and −½ or between +½ and +1, the distribution is moderately skewed.

If skewness is between −½ and +½, the distribution is approximately symmetric.

KURTOSIS:

 This provides information for the degree of peakedness of a data distribution. Peakedness can simply be defined as the measure which data values are concentrated around the mean.

A normal distribution has kurtosis exactly 3 (excess kurtosis exactly 0). Any distribution with kurtosis ≈3 (excess ≈0) is called mesokurtic.
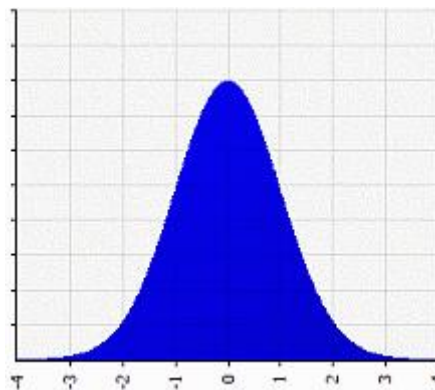


A distribution with kurtosis <3 (excess kurtosis <0) is called platykurtic. Compared to a normal distribution, its tails are shorter and thinner, and often its central peak is lower and broader.
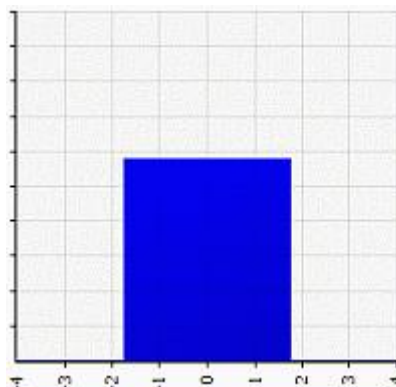
A distribution with kurtosis >3 (excess kurtosis >0) is called leptokurtic. Compared to a normal distribution, its tails are longer and fatter, and often its central peak is higher and sharper.
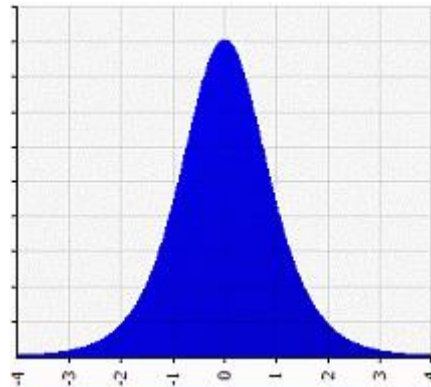


## STANDARD DEVIATION:

This can simply be defined as the level of dispersion of data in relation to the mean. Low standard deviation relates to the data being clustered around the mean while high standard deviation relates to the data being more spread out.
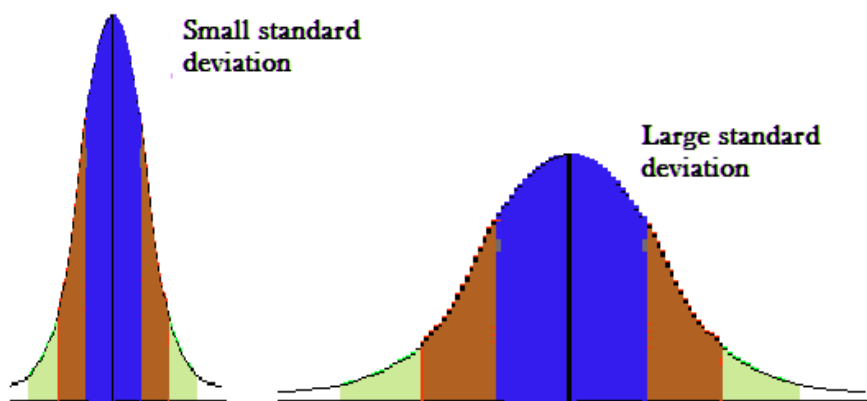


## REGRESSION ANALYSIS:

This concept is basically used to measure the strength of relationship between a dependent and independent variable and also can be used to forecast the future relationship.

Dependent Variable: Also known as the outcome,  is the main factor that you're trying to understand or predict.

Independent Variables: These are the KPIs which impact the variable dependent variable.

For example, we could decide to use the basic income and expenditure. As for you to be able to spend, you need to have an income so in this case, the expenditure is your Dependent variable and Income is your independent variable.

There are three main types of regression namely Linear regression, Multiple regression and Nonlinear regression.
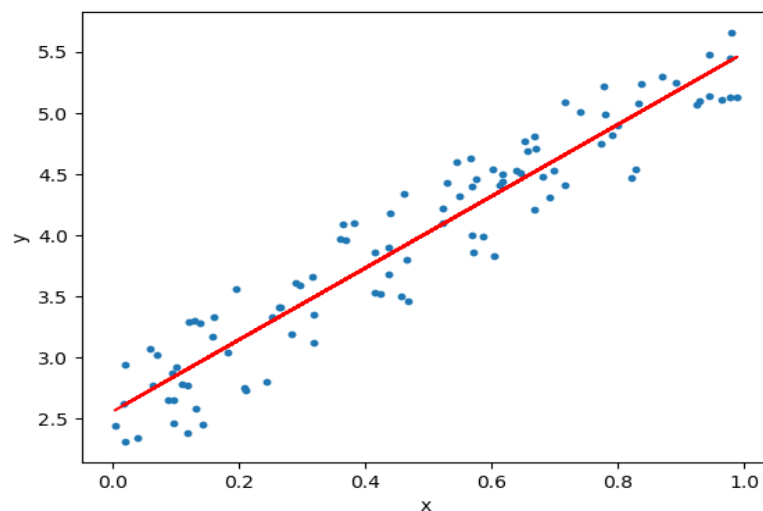


CORRELATION ANALYSIS:

This concept helps to identify the relationship between data of two KPIs collected, a dependent variable and independent variable same as Regression.

A positive correlation result means that both variables increase in relation to each other, while a negative correlation means that as one variable decreases, the other increases. And is measured on a scale of -1 to +1
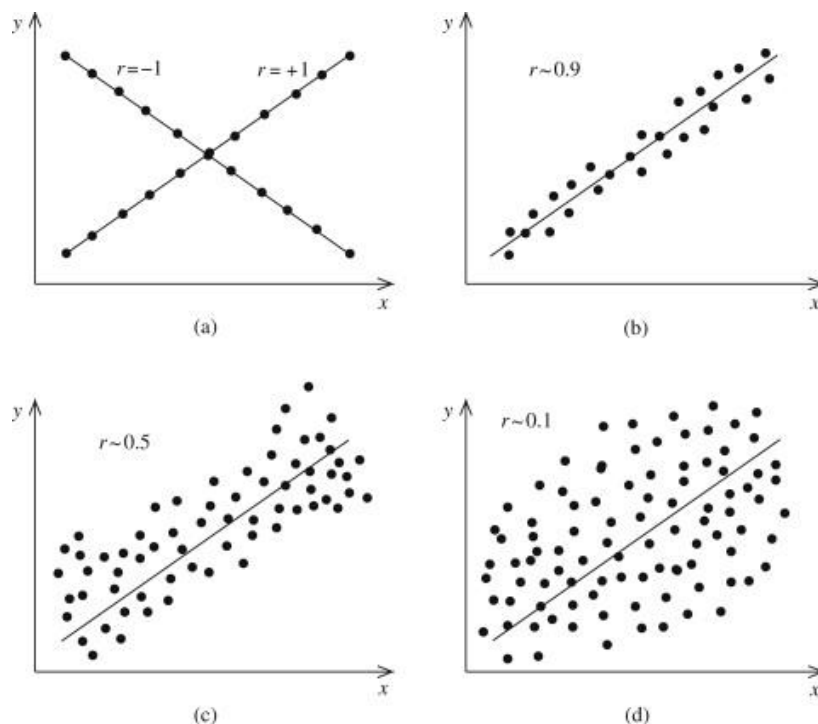


As seen from the image above, a value of 1 meets the follows the lines with no dispersion but with 0.5 and 0.1, there is a level of dispersion and expands more as it goes from 0.5 to 0.1.

## Hypothesis Testing:

This is used to test a hypothesis in relation to a statistical parameter. It is basically verification of an assumption that could help develop a statistical parameter

After analysis of data, you get results that establishes whether the alternative hypothesis stands true or not. When the P-value is less than the significance level, 0.05 the null hypothesis is rejected and the alternative hypothesis fails to be rejected i.e above 0.5

# 3. EXPLORATION OF DATASET

Same as my visualization earlier, I have used the same variables. All variables I picked ranged over 15 years from 2016 to 2020 and are listed below:

1. GDP (current US$)
2. Access to electricity, urban (% of urban population)
3. Access to electricity (% of population)
4. Antiretroviral therapy coverage (% of people living with HIV)
5. Children (0-14) living with HIV
6. Children (ages 0-14) newly infected with HIV
7. Current health expenditure (% of GDP)
8. Incidence of tuberculosis (per 100,000 people)
9. Incidence of malaria (per 1,000 population at risk)
10. International tourism, number of arrivals
11. Newborns protected against tetanus (%)
12. Labor force, female (% of total labor force)
13. Labor force, total
14. Exports of goods and services (% of GDP)
15. Fixed telephone subscriptions
16. Fixed broadband subscriptions
17. Mobile cellular subscriptions
18. Urban population
19. Individuals using the Internet (% of population)

These countries used are listed below:

1. Germany
2. France
3. Spain
4. Netherlands
5. Italy
6. Algeria
7. Tunisia

8. Morocco
9. Egypt
10. Burundi
11. Gambia
12. Lesotho
13. Guinea-Bissau


Before importing, I cleaned the data in Microsoft Excel. By cleaning, I noticed empty rows within the data and I filled them by calculating mean values. After this, I then imported the data into SAS Enterprise Guide and ran the general distribution analysis on the data to check for Extreme Rows and observations. As there was none, I went further to transpose to make better sense of the data as to have the variables I needed as columns so I would be able to run the analysis by Variable, rather than by Country.

# 4. ANALYSIS

## 4.1 DESCRIPTIVE ANALYSIS:

For my descriptive analysis, I ran for all variables of each country to get the full perspective with the below code and the results coming after:

```
proc means data=WORK.TRNSTRANSPOSED mean median kurt skew stddev;
     VAR "GDP (current US$)"n "Access to electricity, urban (%"n
"Access to electricity (% of popu"n "Antiretroviral therapy
coverage"n "Children (0-14) living with HIV"n "Children (ages 0-14)
newly infec"n "Current health expenditure (% of"n
        "Incidence of tuberculosis (per 1"n "Incidence of malaria
(per 1,000"n "International tourism, number of"n "Newborns protected
against tetan"n "Labor force, female (% of total"n "Labor force,
total"n "Exports of goods and services (%"n
        "Fixed telephone subscriptions"n "Fixed broadband
subscriptions"n "Mobile cellular subscriptions"n "Urban population"n
"Individuals using the Internet ("n;
     BY "Country Name"n;
 Title 'Distribution Analysis by Country Name';
run;
```

Country Name=Algeria

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| GDP (current US$) | 1.69457E+11 | 1.70097E+11 | -0.6748315 | 0.0293699 | 29124384896 |
| Access to electricity, urban (% | 99.5663364 | 99.7009735 | -0.7265554 | -0.8066971 | 0.3423549 |
| Access to electricity (% of popu | 99.1300116 | 99 | -0.9888342 | 0.5839373 | 0.2956413 |
| Antiretroviral therapy coverage | 54.8666667 | 56 | -1.6877015 | -0.1055347 | 24.4127211 |
| Children (0-14) living with HIV | 456 | 500 | -0.8567234 | 0.1151989 | 213.4679635 |
| Children (ages 0-14) newly infec | 140 | 100 | -2.0940171 | 0.4550831 | 50.7092553 |
| Current health expenditure (% of | 5.6291312 | 6.0357628 | -0.0676158 | -0.9754577 | 1.0894874 |
| Incidence of tuberculosis (per 1 | 71.7333333 | 73 | 0.5358288 | -0.9402653 | 5.8732647 |
| Incidence of malaria (per 1,000 | 0.002966 | 0 | 7.8414755 | 2.8082813 | 0.0069639 |
| International tourism, number of | 2175066.67 | 2200000 | -1.3382402 | -0.000118272 | 367051.86 |
| Newborns protected against tetan | 87.4 | 90 | -0.9149332 | -0.8495951 | 11.0375981 |
| Labor force, female (% of total | 18.3310671 | 18.8328168 | -1.2658337 | -0.4103451 | 1.7330291 |
| Labor force, total | 11569925.27 | 11584915 | -1.1719457 | -0.2698714 | 817295.8 |
| Exports of goods and services (% | 32.6864599 | 33.2098979 | -1.2178437 | 0.2420557 | 10.3295064 |
| Fixed telephone subscriptions | 3430118.8 | 3132829 | -0.0769667 | 1.0413416 | 671209.99 |
| Fixed broadband subscriptions | 1763642.2 | 1280000 | -1.4170735 | 0.39427 | 1250855.86 |
| Mobile cellular subscriptions | 38087405.53 | 39517045 | -0.6405187 | -0.680129 | 8350176.35 |

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| Urban population | 26735437.93 | 26536380 | -1.226594 | 0.1484345 | 3432109.16 |
| Individuals using the Internet ( | 29.0141484 | 22.5 | -1.2331248 | 0.5176208 | 19.3042922 |

Country Name=Burundi

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| GDP (current US$) | 2300575905 | 2451625333 | -0.7753477 | -0.6312425 | 568595562 |
| Access to electricity, urban (% | 55.1557953 | 52.7109184 | 0.7843845 | -0.572462 | 6.6059884 |
| Access to electricity (% of popu | 7.4482762 | 6.9 | 0.0014875 | 0.5420644 | 2.9052328 |
| Antiretroviral therapy coverage | 42.8666667 | 36 | -1.1918469 | 0.4400587 | 27.697 |
| Children (0-14) living with HIV | 16913.33 | 16000 | -1.1999008 | 0.2839721 | 5432.56 |
| Children (ages 0-14) newly infec | 1649.33 | 1500 | -1.0478601 | 0.6133316 | 710.6985162 |
| Current health expenditure (% of | 8.671833 | 8.1425619 | -1.1534282 | 0.3773889 | 1.461049 |
| Incidence of tuberculosis (per 1 | 133.7333333 | 128 | -0.4572769 | 0.6499946 | 23.4047207 |
| Incidence of malaria (per 1,000 | 207.5233017 | 222.7746531 | -1.0504124 | -0.6968783 | 39.8001437 |
| International tourism, number of | 207333.33 | 205000 | -0.6357012 | 0.3112096 | 53677.17 |
| Newborns protected against tetan | 84.6666667 | 85 | 1.7855067 | -1.1353399 | 7.3452284 |
| Labor force, female (% of total | 52.0926927 | 52.1158569 | 1.982059 | 0.5059322 | 0.1383239 |
| Labor force, total | 4138369.07 | 4085944 | -1.0841056 | 0.1953817 | 593915.32 |
| Exports of goods and services (% | 7.9633409 | 7.6034188 | -1.0146077 | -0.2059395 | 1.6368081 |
| Fixed telephone subscriptions | 24799 | 23409 | -0.9967357 | 0.3879024 | 5334.78 |
| Fixed broadband subscriptions | 1820.73 | 1534 | -1.7667218 | 0.3458928 | 1658.34 |
| Mobile cellular subscriptions | 3279896.27 | 2536831 | -1.6728247 | 0.1863942 | 2439005.47 |
| Urban population | 1128071.07 | 1095417 | -1.0756513 | 0.313001 | 286221.85 |
| Individuals using the Internet ( | 2.1281706 | 1.2642181 | 0.0845385 | 1.2484808 | 1.7169368 |

Country Name=Egypt, Arab Rep.

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| GDP (current US$) | 2.48903E+11 | 2.49713E+11 | -0.6782126 | -0.4177347 | 76477093641 |
| Access to electricity, urban (% | 99.7340763 | 99.8602829 | -0.1078616 | -0.8890667 | 0.2851962 |
| Access to electricity (% of popu | 99.5847392 | 99.8 | 0.3987928 | -1.1168831 | 0.4709195 |
| Antiretroviral therapy coverage | 19.3333333 | 14 | -0.5466628 | 0.8897323 | 12.6923299 |
| Children (0-14) living with HIV | 260 | 200 | -1.6382556 | 0.6124171 | 180.4755623 |
| Children (ages 0-14) newly infec | 106.6666667 | 100 | 15 | 3.8729833 | 25.819889 |
| Current health expenditure (% of | 4.8571413 | 4.9173093 | -0.9110237 | 0.0657465 | 0.4338413 |
| Incidence of tuberculosis (per 1 | 15.9333333 | 16 | -1.1378676 | -0.0258894 | 3.1274514 |
| Incidence of malaria (per 1,000 | 0 | 0 | . | . | 0 |
| International tourism, number of | 10692400 | 11091000 | 0.8238308 | -0.5032783 | 2297054.88 |
| Newborns protected against tetan | 85.4 | 86 | 12.481685 | -3.4420607 | 1.5491933 |
| Labor force, female (% of total | 22.6081397 | 22.9805424 | -0.3351537 | -1.0019719 | 1.182438 |
| Labor force, total | 28130880.47 | 29077255 | -0.3011177 | -0.9725991 | 1915542.74 |
| Exports of goods and services (% | 19.7761375 | 17.5006107 | -0.4975934 | 0.7402567 | 6.9036455 |

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| Fixed telephone subscriptions | 8644708.2 | 8714286 | -1.3091966 | 0.1127914 | 1937994.82 |
| Fixed broadband subscriptions | 3401017.27 | 2676263 | -0.0895465 | 0.881215 | 2763500.4 |
| Mobile cellular subscriptions | 77992507.2 | 94016152 | 0.0143065 | -1.1954082 | 28139394.27 |
| Urban population | 38103249.53 | 37919407 | -1.2328377 | 0.1567834 | 3454279.02 |
| Individuals using the Internet ( | 33.6489108 | 29.4 | 0.4828367 | 0.9308547 | 16.5256615 |

Country Name=France

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| GDP (current US$) | 2.67E+12 | 2.68E+12 | -0.0672976 | -0.5862206 | 1.68079E+11 |
| Access to electricity, urban (% | 100 | 100 | . | . | 0 |
| Access to electricity (% of popu | 100 | 100 | . | . | 0 |
| Antiretroviral therapy coverage | . | . | . | . | . |
| Children (0-14) living with HIV | . | . | . | . | . |
| Children (ages 0-14) newly infec | . | . | . | . | . |
| Current health expenditure (% of | 11.1648766 | 11.2967053 | 0.5965944 | -1.3971901 | 0.4061302 |
| Incidence of tuberculosis (per 1 | 9.2066667 | 9 | 1.4604991 | 1.3896443 | 0.8302036 |
| Incidence of malaria (per 1,000 | . | . | . | . | . |
| International tourism, number of | 200580600 | 203000000 | -1.2234805 | 0.1493093 | 7225120.74 |
| Newborns protected against tetan | . | . | . | . | . |
| Labor force, female (% of total | 47.5463955 | 47.5527074 | 0.275912 | 0.4018998 | 0.4769443 |
| Labor force, total | 29891375 | 30033626 | 0.0238279 | -0.87993 | 442356.82 |
| Exports of goods and services (% | 29.0103661 | 29.2030272 | 0.2011463 | -0.474763 | 1.8832388 |
| Fixed telephone subscriptions | 38259800 | 38805000 | -0.0489583 | -0.8841179 | 2087610.12 |
| Fixed broadband subscriptions | 23842133.33 | 24940000 | -0.4001434 | -0.7037246 | 5403436.38 |
| Mobile cellular subscriptions | 63333540 | 63324000 | -1.040032 | -0.1436041 | 6452047 |
| Urban population | 52093921.13 | 52228059 | -1.2259197 | -0.1875325 | 1725917.51 |
| Individuals using the Internet ( | 76.4038692 | 79.2698113 | 5.7360781 | -2.2178128 | 9.7717235 |

Country Name=Gambia, The

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| GDP (current US$) | 1469182483 | 1450140386 | 0.4194086 | 0.1699523 | 210901276 |
| Access to electricity, urban (% | 69.5915412 | 67.7480011 | -1.1391209 | 0.2341864 | 7.6086236 |
| Access to electricity (% of popu | 50.6955555 | 51.5 | -1.0820726 | -0.2953768 | 7.6831259 |
| Antiretroviral therapy coverage | 17.1333333 | 17 | -1.3596225 | -0.1300961 | 10.2877643 |
| Children (0-14) living with HIV | 2213.33 | 2200 | 1.95671 | -1.1272802 | 83.3809388 |
| Children (ages 0-14) newly infec | 500 | 500 | . | . | 0 |
| Current health expenditure (% of | 3.9341297 | 3.9137487 | -0.6916209 | 0.5417544 | 0.7299767 |
| Incidence of tuberculosis (per 1 | 173.8666667 | 176 | -0.6326 | -0.4854045 | 9.5683008 |
| Incidence of malaria (per 1,000 | 201.3645316 | 236.9934838 | -1.7188309 | -0.3060481 | 104.0177864 |
| International tourism, number of | 292933.33 | 157000 | -1.7058356 | 0.5539074 | 201952.42 |
| Newborns protected against tetan | 91.0666667 | 92 | 11.2802061 | -3.1360076 | 2.6583203 |

| | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| Labor force, female (% of total | 43.5485626 | 43.6193298 | 2.272677 | 0.8636862 | 0.6872708 |
| Labor force, total | 641903.07 | 639157 | -1.3239038 | 0.0293048 | 93946.14 |
| Exports of goods and services (% | 17.4898365 | 18.0329575 | 2.5158948 | -1.2502592 | 3.5588072 |
| Fixed telephone subscriptions | 47386 | 48451 | 0.0573655 | 0.4290853 | 9148.48 |
| Fixed broadband subscriptions | 1971.2 | 500 | -2.101984 | 0.2592076 | 1824.84 |
| Mobile cellular subscriptions | 1945722.13 | 1848854 | -1.2003907 | -0.1827423 | 862390.93 |
| Urban population | 1150968.67 | 1135416 | -1.1714915 | 0.1950712 | 215653.25 |
| Individuals using the Internet ( | 20.7821505 | 14 | 0.1722156 | 1.1594432 | 16.6111211 |

Country Name=Germany

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| GDP (current US$) | 3.60E+12 | 3.68E+12 | 0.2559656 | -0.6524993 | 2.66364E+11 |
| Access to electricity, urban (% | 100 | 100 | . | . | 0 |
| Access to electricity (% of popu | 100 | 100 | . | . | 0 |
| Antiretroviral therapy coverage | 70.7333333 | 71 | -1.1427787 | -0.0372484 | 10.0104707 |
| Children (0-14) living with HIV | . | . | . | . | . |
| Children (ages 0-14) newly infec | . | . | . | . | . |
| Current health expenditure (% of | 10.9805231 | 11.0968647 | -0.2506202 | -0.8579242 | 0.4819564 |
| Incidence of tuberculosis (per 1 | 6.5066667 | 6.1 | -1.0656283 | 0.6861308 | 0.8647598 |
| Incidence of malaria (per 1,000 | . | . | . | . | . |
| International tourism, number of | 31581266.67 | 31545000 | -1.5466622 | 0.0314463 | 5975740.25 |
| Newborns protected against tetan | . | . | . | . | . |
| Labor force, female (% of total | 45.8979187 | 46.0871645 | -1.1395646 | -0.644243 | 0.4764875 |
| Labor force, total | 42459795.07 | 42104556 | -1.1588646 | 0.6393626 | 769957.27 |
| Exports of goods and services (% | 44.6094437 | 45.4186779 | 1.6663018 | -1.2698134 | 2.5380459 |
| Fixed telephone subscriptions | 47858066.67 | 48700000 | -0.8497525 | -0.4474087 | 5037351.66 |
| Fixed broadband subscriptions | 28204720.07 | 28641961 | 0.3689741 | -0.7791513 | 5875267.51 |
| Mobile cellular subscriptions | 99693466 | 100034000 | -1.0702931 | -0.4626472 | 7714658.19 |
| Urban population | 63094461.27 | 62877220 | -0.8988602 | 0.3861899 | 799201.04 |
| Individuals using the Internet ( | 82.7622485 | 84.1652066 | -0.0403747 | -0.6966126 | 4.9894822 |

Country Name=Guinea-Bissau

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| GDP (current US$) | 1065397837 | 1048229629 | -0.8137646 | 0.0932797 | 276355433 |
| Access to electricity, urban (% | 34.92342 | 33.1 | -0.3537907 | 0.2466317 | 12.2473241 |
| Access to electricity (% of popu | 17.7877111 | 16.2006149 | -0.9284366 | 0.314462 | 9.1509642 |
| Antiretroviral therapy coverage | 22.1333333 | 18 | 0.0706885 | 0.7799356 | 16.6727608 |
| Children (0-14) living with HIV | 4126.67 | 4100 | -1.320949 | -0.0914459 | 376.9552203 |
| Children (ages 0-14) newly infec | 674.6666667 | 580 | -1.8804267 | 0.3387485 | 182.1642378 |
| Current health expenditure (% of | 7.1580444 | 7.000968 | 2.6255532 | 1.302386 | 0.9118359 |
| Incidence of tuberculosis (per 1 | 361 | 361 | . | . | 0 |

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| Incidence of malaria (per 1,000 | 100.2892883 | 93.899729 | -1.2852129 | 0.421737 | 23.5337501 |
| International tourism, number of | 34740 | 36100 | -1.3566627 | -0.2973587 | 14643.86 |
| Newborns protected against tetan | 82.6 | 80 | 0.8767154 | 0.3194942 | 6.5224886 |
| Labor force, female (% of total | 47.8474983 | 47.8708377 | -1.478086 | -0.3714524 | 0.0999187 |
| Labor force, total | 679228.93 | 676792 | -1.3596251 | 0.032933 | 83018.48 |
| Exports of goods and services (% | 20.5357173 | 20.1910356 | -1.2503224 | 0.1104247 | 4.99306 |
| Fixed telephone subscriptions | 5341.67 | 5300 | 2.2843746 | 1.1607276 | 554.9265932 |
| Fixed broadband subscriptions | 1149.8 | 1001 | 7.0970581 | 2.431294 | 393.0732756 |
| Mobile cellular subscriptions | 997688.8 | 1049193 | -0.7788925 | 0.0098869 | 505964.48 |
| Urban population | 688308.6 | 680896 | -1.1913256 | 0.1840866 | 108983.47 |
| Individuals using the Internet ( | 10.1571773 | 3.1 | 0.0217397 | 1.1432448 | 10.8613875 |

Country Name=Italy

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| GDP (current US$) | 2.08E+12 | 2.09E+12 | -0.5286984 | 0.2044191 | 1.61354E+11 |
| Access to electricity, urban (% | 100 | 100 | . | . | 0 |
| Access to electricity (% of popu | 100 | 100 | . | . | 0 |
| Antiretroviral therapy coverage | 72.2666667 | 72 | -0.9496188 | 0.1719438 | 11.1064117 |
| Children (0-14) living with HIV | . | . | . | . | . |
| Children (ages 0-14) newly infec | . | . | . | . | . |
| Current health expenditure (% of | 8.7279714 | 8.7754383 | 2.7133668 | -1.5517084 | 0.2203929 |
| Incidence of tuberculosis (per 1 | 6.9466667 | 6.6 | -0.7460449 | 0.2837778 | 0.9869626 |
| Incidence of malaria (per 1,000 | . | . | . | . | . |
| International tourism, number of | 79561773.44 | 76762000 | -1.026632 | 0.4667115 | 9109861.44 |
| Newborns protected against tetan | . | . | . | . | . |
| Labor force, female (% of total | 41.7909479 | 42.1307632 | -1.1112559 | -0.6304633 | 0.8483066 |
| Labor force, total | 25129545.07 | 25080671 | -1.6401273 | 0.1426661 | 618747.04 |
| Exports of goods and services (% | 28.2010315 | 28.6332431 | 0.7218709 | -0.7809213 | 2.4778429 |
| Fixed telephone subscriptions | 21505525.4 | 21098033 | 4.7835167 | 1.8113499 | 1837072.66 |
| Fixed broadband subscriptions | 14019316.47 | 14012976 | -0.3413632 | -0.4308156 | 2731718.59 |
| Mobile cellular subscriptions | 88160333 | 89801000 | -1.0292209 | -0.1176648 | 6285254.37 |
| Urban population | 41294439.33 | 41548775 | -1.5722662 | -0.3424349 | 1109494.35 |
| Individuals using the Internet ( | 57.5512183 | 55.829998 | -0.3342547 | 0.3388423 | 12.3134483 |

Country Name=Lesotho

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| GDP (current US$) | 2175087696 | 2306843529 | -1.4912315 | -0.455909 | 316428719 |
| Access to electricity, urban (% | 58.247867 | 60.650074 | -0.8509931 | -0.3908597 | 13.6818169 |
| Access to electricity (% of popu | 27.3479741 | 26.7335815 | -1.0136272 | 0.4030686 | 12.2198526 |
| Antiretroviral therapy coverage | 45.2 | 42 | -1.0200011 | 0.0821407 | 23.0997835 |
| Children (0-14) living with HIV | 16440 | 17000 | -1.4391479 | -0.249725 | 4732.83 |

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| Children (ages 0-14) newly infec | 1432.67 | 980 | 0.4962269 | 1.2971733 | 1118.12 |
| Current health expenditure (% of | 8.4382935 | 8.5781899 | 0.692699 | -0.2755191 | 1.2861893 |
| Incidence of tuberculosis (per 1 | 947.7333333 | 948 | -1.8817179 | -0.006497 | 244.1508511 |
| Incidence of malaria (per 1,000 | . | . | . | . | . |
| International tourism, number of | 721137.07 | 433000 | -2.1959406 | 0.1443633 | 390500.3 |
| Newborns protected against tetan | 83.2666667 | 83 | 1.5177289 | -0.5628461 | 1.3345233 |
| Labor force, female (% of total | 46.0343241 | 45.6991737 | -1.154217 | 0.6375278 | 0.8150586 |
| Labor force, total | 915811.73 | 902728 | -1.3447077 | 0.5405571 | 34940 |
| Exports of goods and services (% | 45.0052809 | 45.4016068 | -0.4571822 | 0.0505502 | 4.9929454 |
| Fixed telephone subscriptions | 35676.2 | 40570 | -0.8896148 | -0.8694798 | 16210.86 |
| Fixed broadband subscriptions | 2311.14 | 1849 | -1.0292746 | 0.6332522 | 1962.77 |
| Mobile cellular subscriptions | 1427904.6 | 1562648 | -1.2526328 | -0.2401679 | 679225.67 |
| Urban population | 532547.4 | 529466 | -1.1746461 | 0.1564416 | 53806.16 |
| Individuals using the Internet ( | 19.7936487 | 15 | -1.5858663 | 0.4203477 | 16.455892 |

Country Name=Morocco

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| GDP (current US$) | 1.00648E+11 | 1.0137E+11 | 0.626912 | -0.8086374 | 14059341456 |
| Access to electricity, urban (% | 98.0733066 | 98.7 | -1.3353104 | -0.4259568 | 1.6875563 |
| Access to electricity (% of popu | 93.9683869 | 96.5 | -0.1110644 | -1.0175739 | 5.9191357 |
| Antiretroviral therapy coverage | 36.0666667 | 32 | -1.1750202 | 0.4535639 | 22.7548006 |
| Children (0-14) living with HIV | 809.3333333 | 820 | -0.4834957 | -0.733556 | 63.6358689 |
| Children (ages 0-14) newly infec | 180 | 200 | 0.8974359 | -1.6720819 | 41.4039336 |
| Current health expenditure (% of | 5.4476029 | 5.42 | -0.9923574 | -0.0341212 | 0.2836501 |
| Incidence of tuberculosis (per 1 | 99.1333333 | 99 | 0.9062617 | -0.4591394 | 2.2318047 |
| Incidence of malaria (per 1,000 | 0 | 0 | . | . | 0 |
| International tourism, number of | 10155133.33 | 10349000 | -0.4416451 | -0.1611399 | 1804228.01 |
| Newborns protected against tetan | 88.6666667 | 88 | 2.0241758 | 1.1831751 | 1.2909944 |
| Labor force, female (% of total | 25.5650645 | 25.8445293 | -1.6163835 | -0.2439117 | 0.9850515 |
| Labor force, total | 11581137.93 | 11742460 | 0.0053514 | -0.9768143 | 379750.9 |
| Exports of goods and services (% | 34.7161171 | 34.8000405 | 1.8835789 | -0.6120157 | 2.7303114 |
| Fixed telephone subscriptions | 2608288.13 | 2393767 | -0.5833273 | 0.1021766 | 701724.05 |
| Fixed broadband subscriptions | 977119.2 | 843940 | -0.4459189 | 0.7647665 | 535111.66 |
| Mobile cellular subscriptions | 36505779.8 | 41513933 | -0.7282125 | -0.8146352 | 10640971.06 |
| Urban population | 20196855.47 | 20128276 | -1.2460295 | 0.0921381 | 2031060.15 |
| Individuals using the Internet ( | 52.1605812 | 56 | 0.1056136 | -0.359766 | 17.6927794 |

Country Name=Netherlands

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| GDP (current US$) | 8.58178E+11 | 8.68077E+11 | -0.1704555 | -0.6775748 | 60758298452 |
| Access to electricity, urban (% | 100 | 100 | . | . | 0 |

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| Access to electricity (% of popu | 100 | 100 | . | . | 0 |
| Antiretroviral therapy coverage | 72.4666667 | 74 | -1.3223113 | -0.2516885 | 11.3883817 |
| Children (0-14) living with HIV | . | . | . | . | . |
| Children (ages 0-14) newly infec | . | . | . | . | . |
| Current health expenditure (% of | 10.0381827 | 10.1550875 | 0.1986233 | -1.1244381 | 0.5060993 |
| Incidence of tuberculosis (per 1 | 6.04 | 5.9 | -0.0835764 | -0.1326946 | 0.9590174 |
| Incidence of malaria (per 1,000 | . | . | . | . | . |
| International tourism, number of | 13891292.93 | 12783000 | -1.2599632 | 0.5516056 | 3528079.56 |
| Newborns protected against tetan | . | . | . | . | . |
| Labor force, female (% of total | 45.8406672 | 45.9762201 | 2.0288856 | 0.5554623 | 0.70556 |
| Labor force, total | 8962849.53 | 8966340 | 0.1857597 | -0.1735957 | 259586.15 |
| Exports of goods and services (% | 76.3279573 | 79.5033738 | -0.7066274 | -0.6899147 | 6.8642583 |
| Fixed telephone subscriptions | 6781300.93 | 7125486 | 1.6277576 | -1.5541419 | 743186.5 |
| Fixed broadband subscriptions | 6645958 | 6792000 | -0.5306077 | -0.6618406 | 730809.05 |
| Mobile cellular subscriptions | 20108490.87 | 20149000 | 1.6588277 | -0.9131258 | 1114355.7 |
| Urban population | 14935595.2 | 14987369 | -1.082774 | -0.1595461 | 764238.87 |
| Individuals using the Internet ( | 90.6026164 | 91.4199958 | 1.1757733 | -1.2920758 | 2.8932152 |

Country Name=Spain

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| GDP (current US$) | 1.37E+12 | 1.37E+12 | 0.167735 | 0.436488 | 1.1392E+11 |
| Access to electricity, urban (% | 100 | 100 | . | . | 0 |
| Access to electricity (% of popu | 100 | 100 | . | . | 0 |
| Antiretroviral therapy coverage | 73 | 73 | -1.0614363 | -0.2500089 | 10.0356507 |
| Children (0-14) living with HIV | . | . | . | . | . |
| Children (ages 0-14) newly infec | . | . | . | . | . |
| Current health expenditure (% of | 8.8722024 | 9.0673771 | 1.8569993 | -1.7600034 | 0.4508761 |
| Incidence of tuberculosis (per 1 | 13.9466667 | 13 | -1.2406905 | 0.1265991 | 3.9610364 |
| Incidence of malaria (per 1,000 | . | . | . | . | . |
| International tourism, number of | 106253333 | 103231000 | -0.9411929 | 0.6023442 | 11334896.73 |
| Newborns protected against tetan | . | . | . | . | . |
| Labor force, female (% of total | 45.0721063 | 45.8662231 | -0.3963494 | -1.0114173 | 1.5317454 |
| Labor force, total | 23092407.6 | 23116192 | 2.9019093 | -1.524277 | 442293.54 |
| Exports of goods and services (% | 30.4461984 | 31.4552725 | -1.4009076 | -0.4592321 | 4.2173919 |
| Fixed telephone subscriptions | 19763378 | 19639778 | -0.3951633 | 0.6784421 | 387528.7 |
| Fixed broadband subscriptions | 12086724.8 | 12252061 | -0.8405874 | -0.3783546 | 2827549.87 |
| Mobile cellular subscriptions | 51377356 | 51067770 | 0.8348962 | -0.256199 | 2551814.92 |
| Urban population | 36725355 | 36891840 | 0.9676256 | -0.854064 | 1006497.11 |
| Individuals using the Internet ( | 72.7926505 | 71.635 | -0.9283122 | -0.0677997 | 12.957635 |

Country Name=Tunisia

| Variable | Mean | Median | Kurtosis | Skewness | Std Dev |
|---|---|---|---|---|---|
| GDP (current US$) | 44033524752 | 44360775649 | 0.9471604 | -0.7372321 | 4077627631 |
| Access to electricity, urban (% | 99.9030826 | 99.93 | -0.2965922 | -0.9517505 | 0.1059793 |
| Access to electricity (% of popu | 99.6933333 | 99.7 | -1.820711 | 0.1212947 | 0.246306 |
| Antiretroviral therapy coverage | 18.9333333 | 18 | -0.7594867 | 0.1892166 | 7.4590948 |
| Children (0-14) living with HIV | 100 | 100 | . | . | 0 |
| Children (ages 0-14) newly infec | 100 | 100 | . | . | 0 |
| Current health expenditure (% of | 6.5192764 | 6.8073964 | -1.4415084 | -0.4798371 | 0.7610328 |
| Incidence of tuberculosis (per 1 | 32.8666667 | 35 | -1.3924018 | -0.6072513 | 4.4700059 |
| Incidence of malaria (per 1,000 | . | . | . | . | . |
| International tourism, number of | 7158000 | 7052000 | 0.1387988 | 0.2091139 | 1083307.44 |
| Newborns protected against tetan | 95.6666667 | 96 | 13.7691184 | -3.6271041 | 1.8771813 |
| Labor force, female (% of total | 27.1769244 | 27.1093015 | -1.3151974 | 0.1212775 | 0.485252 |
| Labor force, total | 3894750 | 3962506 | -0.5367565 | -0.8365727 | 203399.04 |
| Exports of goods and services (% | 45.4262816 | 45.6782556 | 1.2078219 | 0.4496696 | 4.4369447 |
| Fixed telephone subscriptions | 1197163.93 | 1239074 | -0.5711286 | 0.1296768 | 177495.08 |
| Fixed broadband subscriptions | 596856.6 | 559358 | 0.0523938 | 0.5571458 | 367052.09 |
| Mobile cellular subscriptions | 12301875.8 | 12843889 | -0.8400783 | -0.8179706 | 2703064.09 |
| Urban population | 7418249.93 | 7392693 | -1.1380095 | 0.1312801 | 487989.64 |
| Individuals using the Internet ( | 43.4608108 | 43.8 | -0.3546401 | -0.1537282 | 16.7536671 |

With the data being excessively much, I would be picking four countries, based on their GDP and go deeper with the most important statistical parameters, two rich countries and two poor countries namely: Burundi, Gambia, France and Germany with the below code:

```
proc means data=WORK.TRNSTRANSPOSED Mean Median skew kurt stddev;
     VAR "GDP (current US$)"n;
     BY "Country Name"n;
     WHERE 'Country Name'n = 'Burundi'OR 'Country Name'n =
'Lesotho' OR 'Country Name'n = 'France' OR 'Country Name'n =
'Germany';
 Title 'Distribution Analysis by Country Name';
run;
```

Country Name=Burundi

| Analysis Variable : GDP (current US$) |
|---|

| Mean | Median | Skewness | Kurtosis | Std Dev |
|---|---|---|---|---|
| 2300575905 | 2451625333 | -0.6312425 | -0.7753477 | 568595562 |

Country Name=France

| Analysis Variable : GDP (current US$) | | | | |
|---|---|---|---|---|
| Mean | Median | Skewness | Kurtosis | Std Dev |
| 2.67E+12 | 2.68E+12 | -0.5862206 | -0.0672976 | 1.68079E+11 |

Country Name=Germany

| Analysis Variable : GDP (current US$) | | | | |
|---|---|---|---|---|
| Mean | Median | Skewness | Kurtosis | Std Dev |
| 3.60E+12 | 3.68E+12 | -0.6524993 | 0.2559656 | 2.66364E+11 |

Country Name=Lesotho

| Analysis Variable : GDP (current US$) | | | | |
|---|---|---|---|---|
| Mean | Median | Skewness | Kurtosis | Std Dev |
| 2175087696 | 2306843529 | -0.455909 | -1.4912315 | 316428719 |

Burundi can be seen to have an average of GDP of $2.3b dollars with a skewness of and a -0.6 skewness meaning its moderately skewed to the left and a kurtosis of -0.7 making it platykurtic.

Lesotho can also be seen to have an average of $2.1b in GDP, a skewness of -0.4 indicating moderate skewness to the left as well and a kurtosis of -1.49 which is also platykurtic.

Germany comes with a mean of $3.6 in GDP and a skewness of -0.65 which indicates moderate skewness to the left as well but a positive kurtosis of 0.25 which is also platykurtic.

France also has a high mean of $2.6T, skewness of -0.58 and a negative kurtosis of -0.06 indicating a platykurtic one as well.

From the above data, it can easily be seen that all countries are moderately skewed and platykurtic indicating its more uniform than it is being leptokurtic as they are no way close to 3 , Germany being the richest country by its mean and Lesotho is the poorest.

## 4.2 CORRELATION ANALYSIS:

As I have done earlier with the Descriptive Statistics, I ran the analysis with GDP as my dependent variable in correlation with International Tourism, Female Labor Force, Exports of goods and services and Individuals using the Internet for all the countries with the code below:

```
ods graphics on;
proc corr data=WORK.TRNSTRANSPOSED rank
      PLOTS=(SCATTER MATRIX);
 var 'gdp (current us$)'n;
 with "Current health expenditure (% of"n "International tourism,
number of"n "Labor force, female (% of total"n "Exports of goods and
services (%"n "Individuals using the Internet ("n;
 by 'Country Name'n;
title "Correlations of key KPIs with GDP";
run;
```
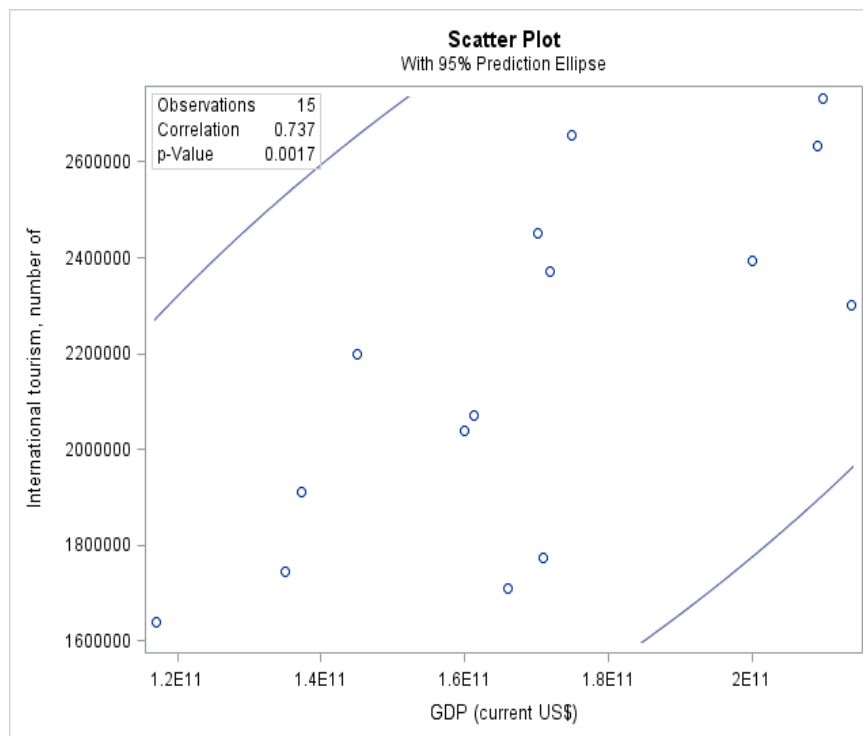
I then filtered to another set of four countries(Algeria, Morocco, Italy, Spain) for my explanation using the following code:

```
ods graphics on;
proc corr data=WORK.TRNSTRANSPOSED rank
      PLOTS=(SCATTER MATRIX);
 var 'gdp (current us$)'n;
 with "Current health expenditure (% of"n "International tourism,
number of"n "Labor force, female (% of total"n "Exports of goods and
services (%"n "Individuals using the Internet ("n;
 by 'Country Name'n;
      WHERE 'Country Name'n = 'Algeria'OR 'Country Name'n =
'Morocco' OR 'Country Name'n = 'Italy' OR 'Country Name'n = 'Spain';
title "Correlations of key KPIs with GDP";
run;
```

Below are my results and necessary scatter plots to go with it

Country Name=Algeria

| Pearson Correlation Coefficients, N = 15 Prob > \|r\| under H0: Rho=0 | |
|---|---|
| Current health expenditure (% of | GDP (current US$) 0.51012 0.052 |
| International tourism, number of | GDP (current US$) 0.73696 0.0017 |
| #Labor force, female (% of total | GDP (current US$) 0.41116 0.1279 |
| Exports of goods and services (% | GDP (current US$) -0.17535 0.5319 |
| Individuals using the Internet ( | GDP (current US$) 0.07084 0.8019 |



**Scatter Plot**
With 95% Prediction Ellipse

| Observations | 15 |
| Correlation | 0.737 |
| p-Value | 0.0017 |

For Algeria, I have decided to show the Scatter Plot of the correlation based on GDP and International Tourism as it has the most positive correlation of 0.73 and also a p-value of 0.0017 which makes the correlation very significant as it is very much below 0.05 based on the prediction with 95% prediction ellipse.

Country Name=Italy

| Pearson Correlation Coefficients, N = 15 Prob > \|r\| under H0: Rho=0 | |
|---|---|
| Current health expenditure (% of | GDP (current US$) -0.24333 0.3822 |
| International tourism, number of | GDP (current US$) -0.50217 0.0564 |
| Labor force, female (% of total | GDP (current US$) -0.47142 0.0761 |
| Exports of goods and services (% | GDP (current US$) -0.4665 0.0796 |
| Individuals using the Internet ( | GDP (current US$) -0.45815 0.0859 |

**Scatter Plot**
With 95% Prediction Ellipse
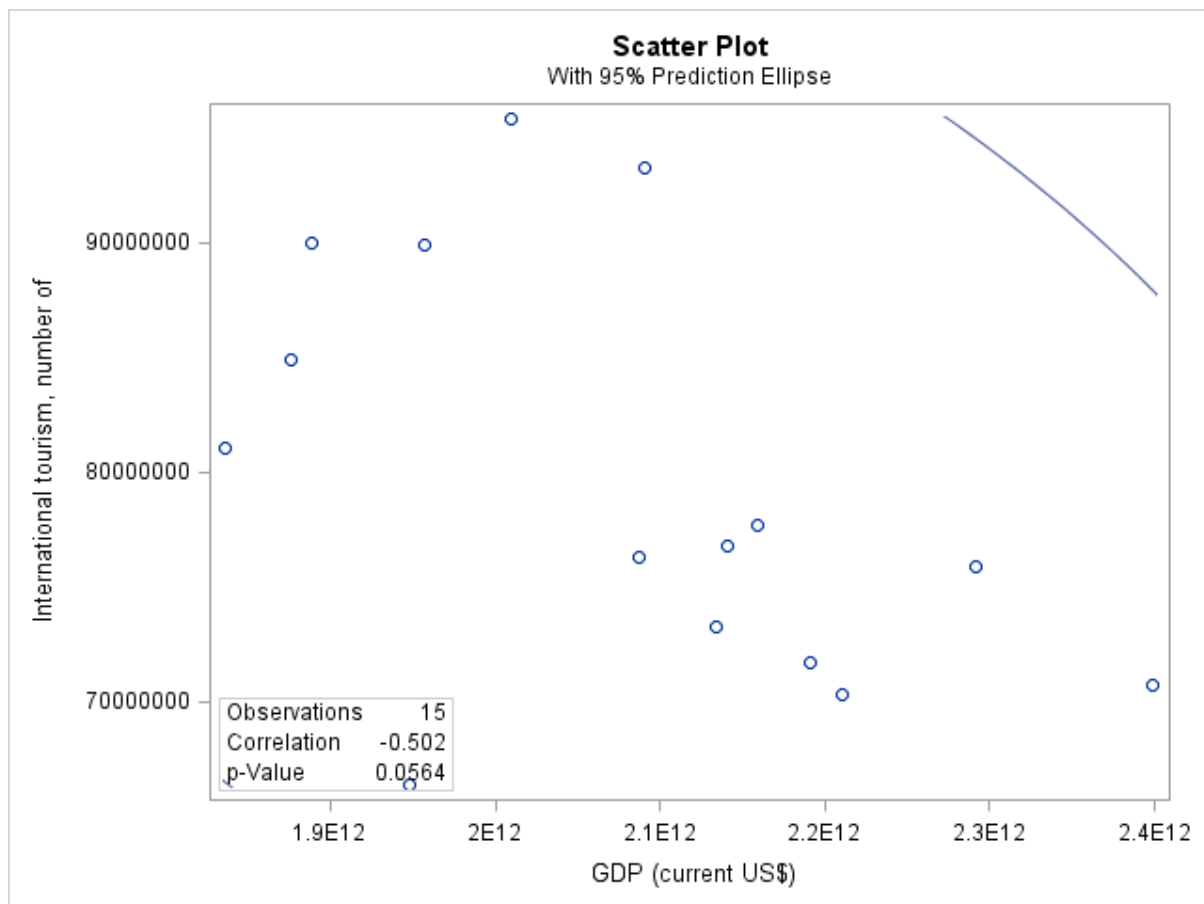
For Italy, I have decided to go with International Tourism and GDP once more as it provides the most negative correlation of 0.50 which means that there was a decline in GDP as well as International Tourism. It also shows a significance p-value of 0.0564. Looking at this data, one can deduce that the GDP of Italy was decreasing steadily and as this the health expenditure reduced as well as there was less money to spend. This also mirrored on its Exports of Goods and Services as it reduced, the GDP was bound to reduce as well.

| Pearson Correlation Coefficients, N = 15 | |
| --- | --- |
| Prob > \|r\| under H0: Rho=0 | |
| Current health expenditure (% of | GDP (current US$) |
| | 0.14034 |
| | 0.6179 |
| International tourism, number of | GDP (current US$) |
| | 0.95685 |
| | <.0001 |
| Labor force, female (% of total | GDP (current US$) |
| | -0.82272 |
| | 0.0002 |
| Exports of goods and services (% | GDP (current US$) |
| | 0.5608 |
| | 0.0296 |
| Individuals using the Internet ( | GDP (current US$) |
| | 0.91838 |
| | <.0001 |



**Scatter Plot**
With 95% Prediction Ellipse

Observations 15
Correlation -0.823
p-Value 0.0002

**Scatter Plot**
With 95% Prediction Ellipse

| Observations | 15 |
| Correlation | 0.9184 |
| p-Value | <.0001 |

*Individuals using the Internet (* — y-axis
*GDP (current US$)* — x-axis
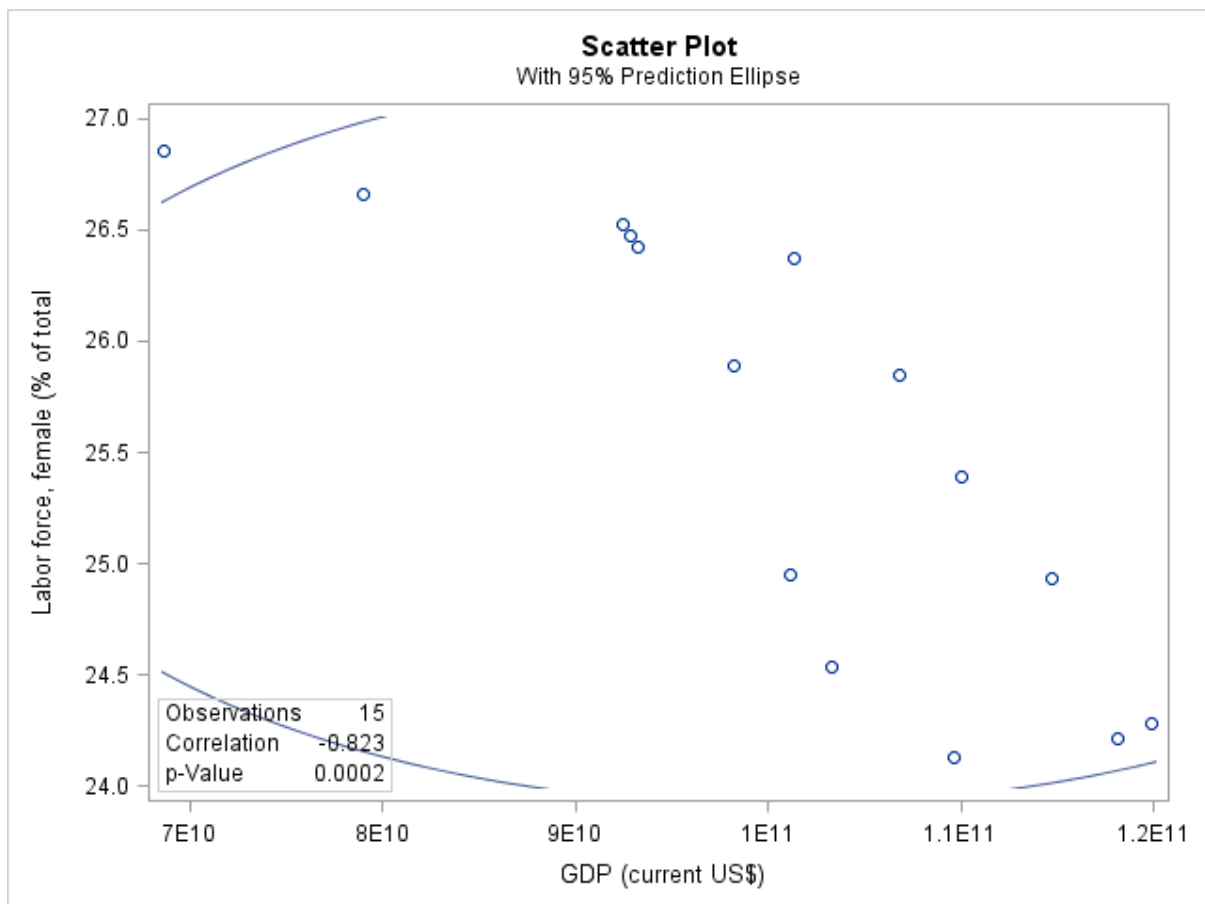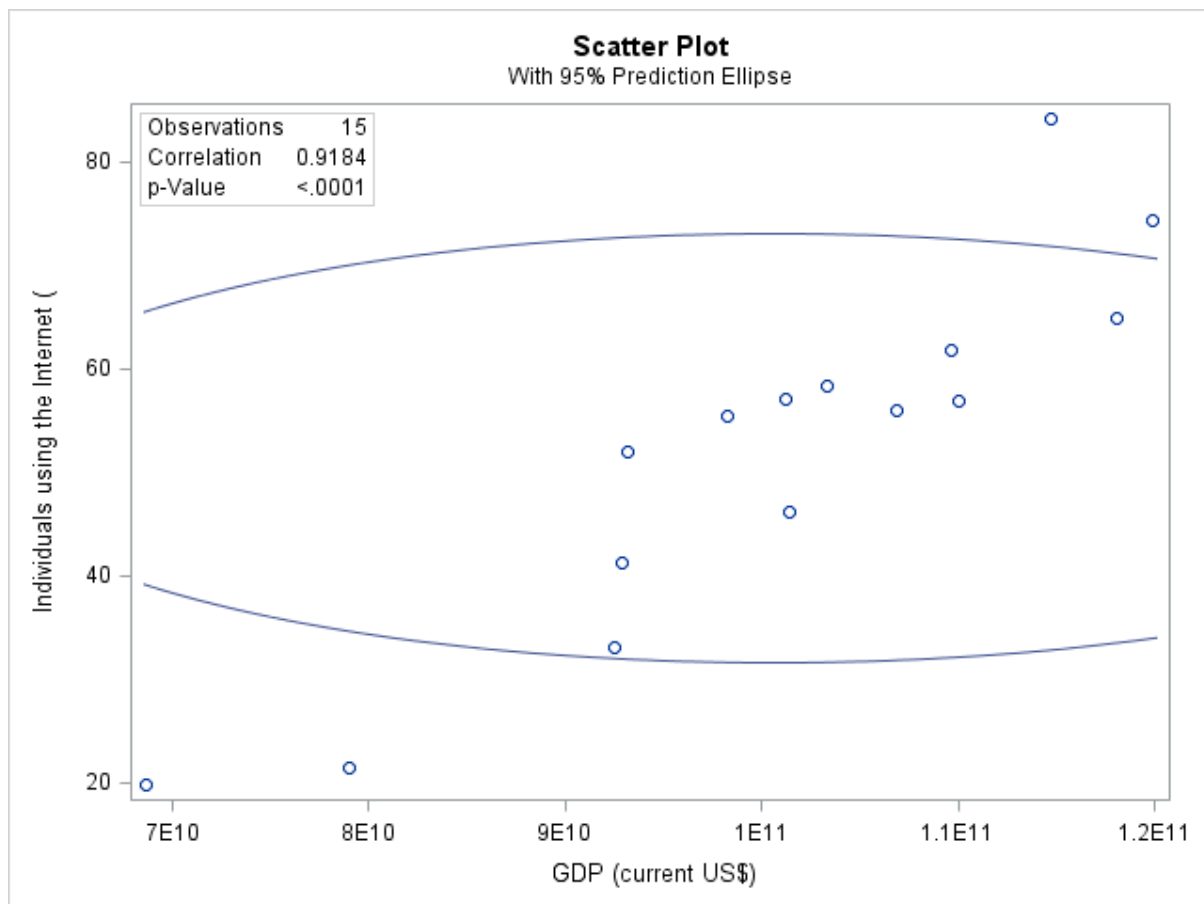
Looking at Morocco, I can see a correlation value of -0.823 which shows a negative correlation between the female labor force and GDP, with a p-value of 0.002 indicating a strong significance between both variables. Also with Individuals using there internet, there was a 0.91 correlation indicating a positive relationship between them which could suggest that as Moroccans got more access to the internet, they probably digitalized their businesses to generate more profit, thereby boosting the economy's GDP.

| Pearson Correlation Coefficients, N = 15 | |
| --- | --- |
| Prob > \|r\| under H0: Rho=0 | |
| Current health expenditure (% of | GDP (current US$) |
| | -0.18709 |
| | 0.5044 |
| International tourism, number of | GDP (current US$) |
| | -0.34728 |
| | 0.2047 |
| Labor force, female (% of total | GDP (current US$) |
| | -0.45674 |
| | 0.087 |
| Exports of goods and services (% | GDP (current US$) |
| | -0.50048 |
| | 0.0574 |
| Individuals using the Internet ( | GDP (current US$) |
| | -0.39974 |
| | 0.1399 |



Scatter Plot
With 95% Prediction Ellipse

| Observations | 15 |
| Correlation | -0.5 |
| p-Value | 0.0574 |
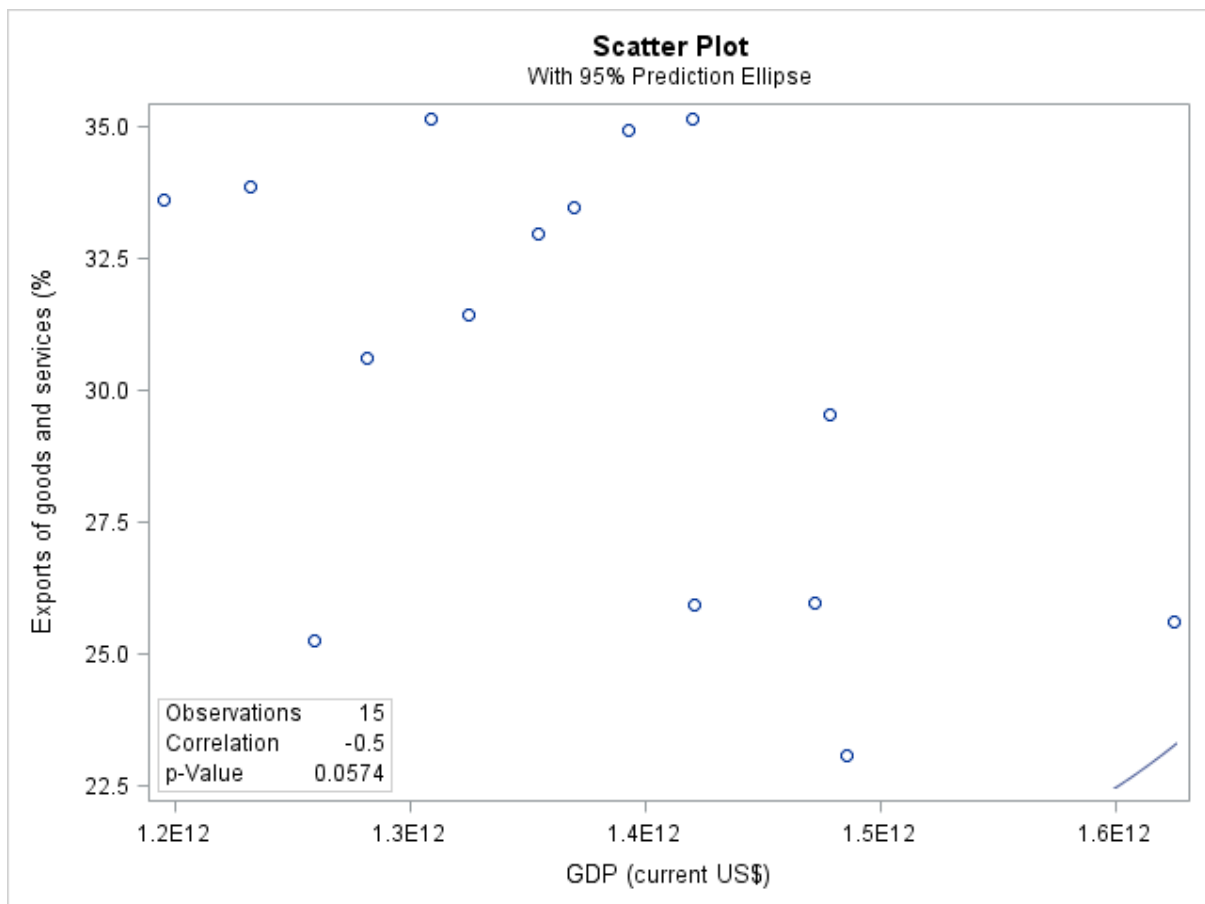
For Spain, I'll be looking at its Exports of Goods and Services in correlation with GDP as it's the most negatively correlated with a value of -0.5 and has p-value closest to 0.05

## 4.3 REGRESSION ANALYSIS:

For this analysis, I have decided to go with Access to Electricity of total people against Access to electricity of urban population as I believe Urban population has more access to electricity than the rural regions so they're the main component of the total people with access to Electricity. With the below code, I have picked four countries namely Algeria, Burundi, Lesotho and Morocco which are all African countries.

Country Name=Algeria

| Number of Observations Read | 15 |
|---|---|
| Number of Observations Used | 15 |

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 1 | 0.44235 | 0.44235 | 7.36 | 0.0177 |
| Error | 13 | 0.7813 | 0.0601 | | |
| Corrected Total | 14 | 1.22365 | | | |

| Root MSE | 0.24515 | R-Square | 0.3615 |
|---|---|---|---|
| Dependent Mean | 99.13001 | Adj R-Sq | 0.3124 |
| Coeff Var | 0.2473 | | |

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
| Intercept | 1 | 47.43401 | 19.0551 | 2.49 | 0.0271 |
| Access to electricity, urban (% | 1 | 0.51921 | 0.19138 | 2.71 | 0.0177 |

As seen, there's an R-Square value of 0.36% which means that 36% of the variation between both variations can be explained by the sample, there's also a p-value of 0.0177 which shows the level of significance.

Country Name=Burundi

| Number of Observations Read | 15 |
|---|---|
| Number of Observations Used | 15 |

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 1 | 80.12511 | 80.12511 | 27.38 | 0.0002 |
| Error | 13 | 38.04017 | 2.92617 | | |
| Corrected Total | 14 | 118.16529 | | | |

| Root MSE | 1.7106 | R-Square | 0.6781 |
|---|---|---|---|
| Dependent Mean | 7.44828 | Adj R-Sq | 0.6533 |
| Coeff Var | 22.96645 | | |

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > |t| |
| Intercept | 1 | -12.52612 | 3.84261 | -3.26 | 0.0062 |
| Access to electricity, urban (% | 1 | 0.36215 | 0.06921 | 5.23 | 0.0002 |

For Burundi, an R-Square value of 0.67 can be seen, which means that 67% of the variation can be defined by the data available and a p-value of 0.0002 indicates a high level of significance between both variable selected.

Country Name=Lesotho

| Number of Observations Read | 15 |
|---|---|
| Number of Observations Used | 15 |

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 1 | 1823.1331 | 1823.1331 | 88.63 | <.0001 |
| Error | 13 | 267.41406 | 20.57031 | | |
| Corrected Total | 14 | 2090.54716 | | | |

| Root MSE | 4.53545 | R-Square | 0.8721 |
|---|---|---|---|
| Dependent Mean | 27.34797 | Adj R-Sq | 0.8622 |
| Coeff Var | 16.58423 | | |

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
| Intercept | 1 | -21.2347 | 5.29171 | -4.01 | 0.0015 |
| Access to electricity, urban (% | 1 | 0.83407 | 0.0886 | 9.41 | <.0001 |

For Lesotho, R-square can be seen as 0.87 meaning 87% of the variation can be explained by the data. Also, the p-value being 0.0001 indicates a high level of significance.

Country Name=Morocco

| Number of Observations Read | 15 |
|---|---|
| Number of Observations Used | 15 |

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 1 | 390.0258 | 390.0258 | 50.46 | <.0001 |
| Error | 13 | 100.48054 | 7.72927 | | |
| Corrected Total | 14 | 490.50634 | | | |

| Root MSE | 2.78016 | R-Square | 0.7951 |
|---|---|---|---|
| Dependent Mean | 93.96839 | Adj R-Sq | 0.7794 |
| Coeff Var | 2.95861 | | |

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > |t| |
| Intercept | 1 | -212.775 | 43.18748 | -4.93 | 0.0003 |
| Access to electricity, urban (% | 1 | 3.12769 | 0.4403 | 7.1 | <.0001 |

## 4.4 Hypothesis Testing:

For this part, I have decided to go with the one sample t test with the GDP of all countries. The One Sample t Test examines whether the mean of the GDP is statistically different from the hypothesized value(Ho) of 2300575905, which happens to be the mean GDP of Burundi.

```
ods graphics;
/*Calculating one sample T Test testing if the mean GDP is
2300575905*/
proc ttest data=work.trnstransposed side=u
plots(shownull)=interval
H0=2300575905;
var 'gdp (current us$)'n;
title "One-Sample t-test testing whether mean GDP across 10
countries is 2300575905";
by 'Country Name'n;
run;
```
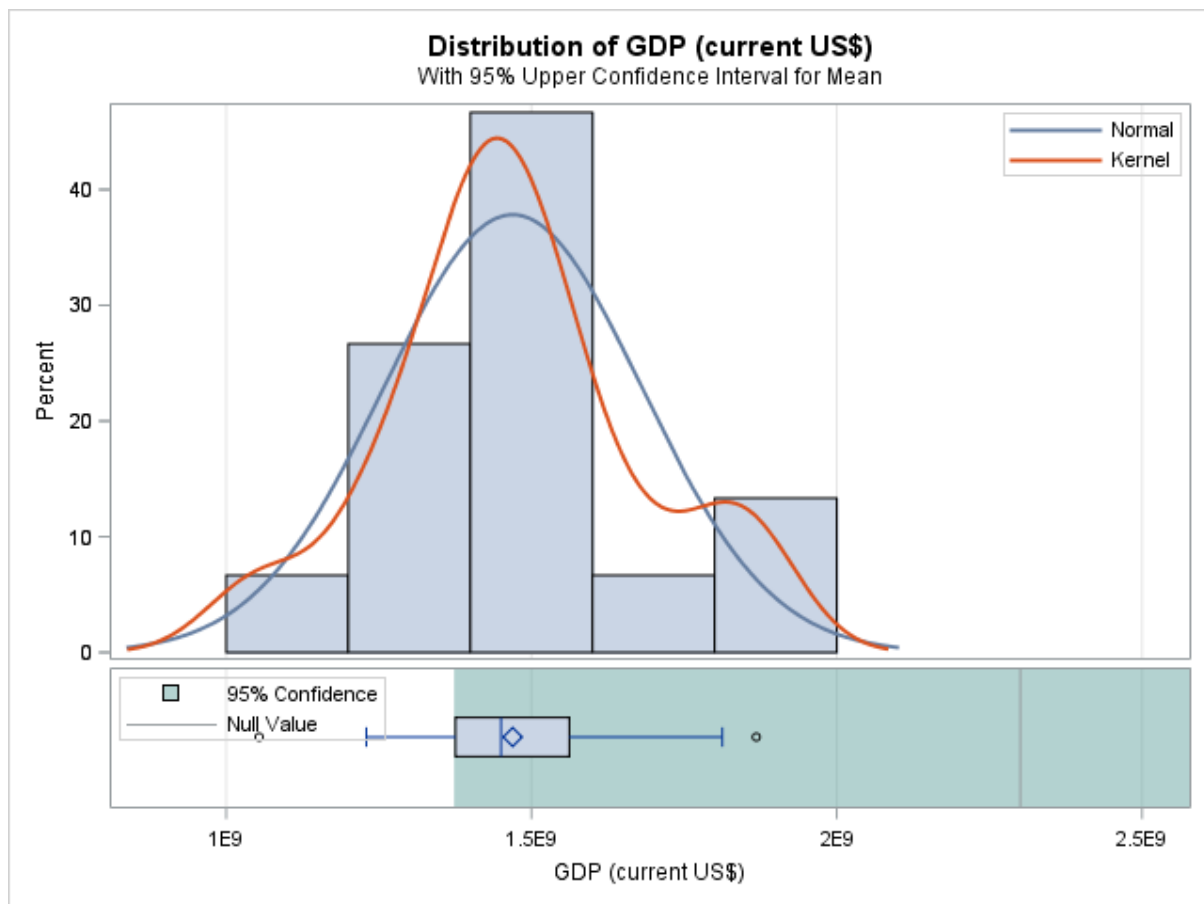
I was able to get the results but I shall be showing results from Gambia and Italy to make for one poor country and one rich country.

Country Name=Gambia, The

| N | Mean | Std Dev | Std Err | Minimum | Maximum |
|---|------|---------|---------|---------|---------|
| 15 | 1.47E+09 | 2.11E+08 | 54454475 | 1.05E+09 | 1.87E+09 |

| Mean | 95% CL Mean | | Std Dev | 95% CL Std Dev | |
|------|-------------|--|---------|----------------|--|
| 1.47E+09 | 1.37E+09 | Infty | 2.11E+08 | 1.54E+08 | 3.33E+08 |

| DF | t Value | Pr > t |
|----|---------|--------|
| 14 | -15.27 | 1 |

Distribution of GDP (current US$)
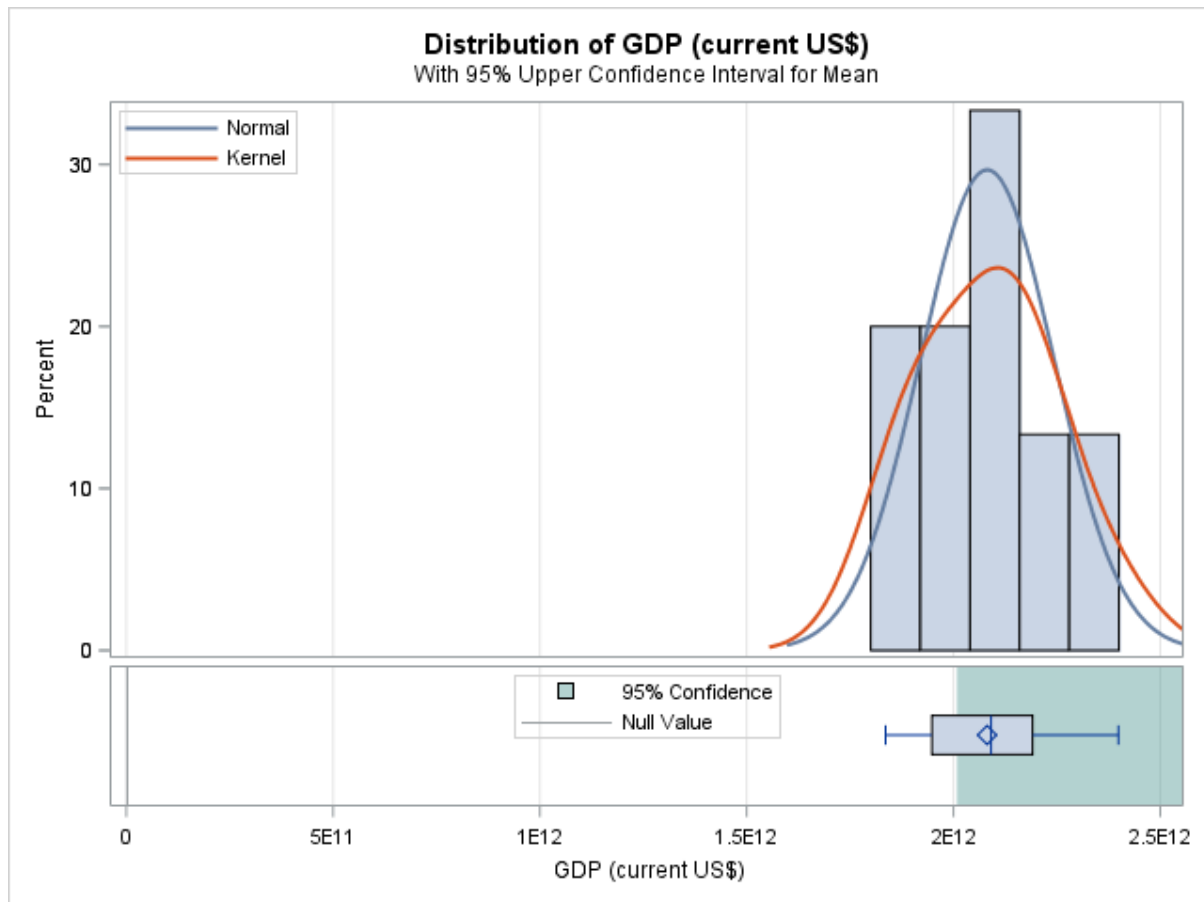With 95% Upper Confidence Interval for Mean

Based on the above data, it can be seen that the p value exceeds the signifcance value (0.05), therefore we fail to reject the null hypothesis that the GDP over time is less than the assumed number.

Country Name=Italy

| N | Mean | Std Dev | Std Err | Minimum | Maximum |
|---|---|---|---|---|---|
| 15 | 2.08E+12 | 1.61E+11 | 4.17E+10 | 1.84E+12 | 2.40E+12 |

| Mean | 95% CL Mean | | Std Dev | 95% CL Std Dev | |
|---|---|---|---|---|---|
| 2.08E+12 | 2.01E+12 | Infty | 1.61E+11 | 1.18E+11 | 2.55E+11 |

| DF | t Value | Pr > t |
|---|---|---|
| 14 | 49.9 | <.0001 |

**Distribution of GDP (current US$)**
With 95% Upper Confidence Interval for Mean

Based on the table above, we can reject the null hypothesis as the p value is below 0.05 as the GDP is higher than the assumed number.

## 4. DISCUSSION AND CONCLUSION

After all analysis carried out, I was able to make many inferences from the Descriptive Analysis, Correlation analysis, Regression Analysis and Hypothesis analysis.

1. **Descriptive analysis**: With this, I was able to find the Minimum, Maximum Mean, Median, Standard Deviation, Skewness and Kurtosis. I also used it to find extreme observations for outliers but was not included in this report. In general, these were my observations:

   - Most of the data was moderately skewed to the left and right, with just one or two variable skewed to both sides

- A little percentage of Kurtosis values had values above 3

2. **Correlation Analysis**: With this, I was able to establish relationships between the effects of exports of good and services, female labor, number of Invidiaul tourists and number of indivduals using the internet on GDP. Also, the effect health expenditure as a measure of the GDP of some countries.

- I was able to use the p-value to evaluate the significance between the variables, with some being above 0.05 and some below 0.05 which isn't surprisng as no model is perfect. From my sample cut out of the original large size, I found the number of international tourists to be the most signifcant as it had the least above 0.05

- I was also able to spot Female Labor Force as the most negative correlated as this occurred three of four times

- I was also able to identify International Tourism as the most correlated with GDP and this makes sense as this usally generates a huge chunk. It went as high as a 95% correlation for Morocco.

3. **Regression Analysis**: Also a relationship parameter, I was also able to find a relationship between Access to Electricity for Urban people compared with Access to Electricity for total population. My observations go thus:

    - In Morocco, for every 3.12% increase in Access to Electricity for the urban population, the Access to electricity of total population increased by 1.

    - In Lesotho, for every 0.18% increase in Access to Electricity for the urban population, the Access to electricity of total population increased by 1.

    - In Burundi, for every 0.36% increase in Access to Electricity for the urban population, the Access to electricity of total population increased 1.

    - In Algeria, for every 0.51% increase in Access to Electricity for the urban population, the Access to electricity of total population increased 1.

- There were pretty much high values for R-Square asides for Algeria with 0.367, indicating only 36.7% of the data could be explained by the variation and Lesotho being the highest with 0.872 indicating 87.2% of the variation could be xplained by the data.

4. **Hypothesis Testing**: Based on my one sample t-test, Italy rejected the null hypothessis while Gambia failed to reject the hypothesis. I also decided to take an extra look at another African country, Egypt and got the following results:

Country Name=Egypt, Arab Rep.

| N | Mean | Std Dev | Std Err | Minimum | Maximum |
|---|---|---|---|---|---|
| 15 | 2.49E+11 | 7.65E+10 | 1.98E+10 | 1.07E+11 | 3.65E+11 |

| Mean | 95% CL Mean | | Std Dev | 95% CL Std Dev | |
|---|---|---|---|---|---|
| 2.49E+11 | 2.14E+11 | Infty | 7.65E+10 | 5.60E+10 | 1.21E+11 |

| DF | t Value | Pr > t |
|---|---|---|
| 14 | 12.49 | <.0001 |

Once again, the hypothesis was rejected by Egypt as well as the GDP was higher than the assumed

**REFERENCES:**

Computing, S. (2020). "How to Build Data Visualization Dashboards | Sigma Computing." Retrieved 2020/01/03, from https://www.sigmacomputing.com/blog/how-to-build-data-visualization-dashboards/.

Data, O. W. i. (2020). "Energy." Retrieved 2022/04/01.

Destinationsinternational.org (2018). "World Travel & Tourism Economic Impact Report | Destinations International." Retrieved 2022/01/04, from https://destinationsinternational.org/industry-reports/world-travel-tourism-economic-impact-report.

Marum (2022). "Care of children with HIV infection and AIDS in Africa." from https://pubmed.ncbi.nlm.nih.gov/9416374/.

Raghupathi, V. (2020). "Healthcare Expenditure and Economic Performance: Insights From the United States Data." Retrieved 2022/01/04, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7237575/.

Unaids.org (2021). "Global HIV & AIDS statistics — Fact sheet." 2021/01/04, from https://www.unaids.org/en/resources/fact-sheet.