**DATA VISUALIZATION FINAL REPORT**

**Subash Chandra Biswal (U77884251)**

***Population and its impact on GDP, Life Expectancy, and Fertility***

**Introduction**

Population growth has been a concern for all. This is because all the resources we need are limited and population growth raises concerns about the sufficient supply of resources for the overall wellbeing of the world.

In 1984, there were approximately 3.6 billion people in developing countries. This was growing at a rate of 2.00 percent annually. At this rate, it was estimated that the population would be doubled in 35 years. This estimation turns out true and now in 2022, we have approximately 8 billion people on earth and the population of developing countries will reach 6.8 billion by 2025.

But with population growth, the fertility rate has declined, and we have achieved unprecedented per capita income, life expectancy, education, and hence GDP.

It’s very much interesting and necessary to analyze the interlinked relationships between population growth, fertility rate, life expectancy, and GDP as these are the primary social and economic parameters of the status of the society.

**Motivation**

It is obvious that the fertility rate navigates population growth, but the relationship is complex and interlinked with other parameters. As the population grows Government policies are imposed to control the fertility rate. With an improved life expectancy rate, families also wish to have fewer children. With the growing population, the demand for natural resources increases the cost of living and also the demand for human resources increases the per capita income and somehow motivates families to have more children. The GDP also plays a major role in improved life and hence influences life expectancy, and education hence it influences fertility and population growth.

It's interesting to analyze these complex relationships and find out if there can be any other parameter influencing all these socio-economic indicators. These findings can be instrumental in planning future policies by Government agencies and institutions.

**Research Questions**

1. How populations of countries across the world have been developing over the 25 years? Does it impact fertility rate or life expectancy? We need to know the overall trend in fertility, life expectancy, and population. We should be able to drill down into individual countries.
2. How population growth has impacted the GDP across the world. We need to know if it helps in GDP growth or not. We should be able to drill down to the individual country to visualize the trends.

**Methodology**

* The data for the project is sourced from ***data.worldbank.org***.
* There are five files retrieved in excel format.
* Country Metadata
* Population Metadata
* Life Expectancy Rate Metadata
* Fertility Rate Metadata
* GDP data
* The files had 214 rows for 214 countries.
* The country metadata file has columns for country code, country name, and region.
* Other 4 files had 25 columns for years (1997-2021) other than country code and country name.
* With Tableau Prep the files are transformed to convert the years’ column to row data with the name “Year”.
* So final excel files after transformation with Tableau Prep are as follows.

|  |  |  |  |
| --- | --- | --- | --- |
| **File Name** | **No. Of Rows** | **No. Of Columns** | **Column Names** |
| Country Metadata | 214 | 3 | Country Code  Country Name  Region |
| Fertility\_extract | 5112 | 4 | Country Code  Country Name  Year  Fertility Rate |
| Life Expectancy\_extract | 5112 | 4 | Country Code  Country Name  Year  Life Expectancy Rate |
| Population\_extract | 5112 | 4 | Country Code  Country Name  Year  Population |
| GDP\_extract | 5112 | 4 | Country Code  Country Name  Year  GDP |

* The columns represent the following information.
* **Country Code**: It represents the 3-letter country code and is unique for each country.
* **Country Name**: It represents the country names.
* **Region**: It represents the region under which the country falls. There are 7 regions.

**East Asia & Pacific**

**Europe & Central Asia**

**Latin America & Caribbean**

**Middle East & North Africa**

**North America**

**South Asia**

**Sub-Saharan Africa**

* **Year**: Represents the year (1997-2021)
* **Fertility Rate**: Rate of fertility for a country for a specific year.
* **Life Expectancy Rate**: Life expectancy rate for a country for a specific year.
* **GDP**: GDP of a country for a specific year.
* **Population**: Population of a country for a specific year.
* The visualizations and analysis are done using Tableau.
* The files are linked using the below relationships.

Diagram

Description automatically generated

|  |  |  |
| --- | --- | --- |
| **File Name** | **Linked File Name** | **Linked Columns** |
| Population\_extract | Country Metadata | Country Code |
| Population\_extract | Life Expectancy\_extract | Country code  Year |
| Population\_extract | Fertility\_extract | Country code  Year |
| Population\_extract | GDP\_extract | Country code  Year |

* The bar graph Y scale is converted to a logarithmic scale to get better visualizations as GDP is very large compared to population, and life expectancy is also very large compared to fertility rate.
* A hierarchy is created for region and country to drill down the graph to country level.

**Analysis**

**Visualization 1**:

Map

Description automatically generated

This visualization is created by plotting the Median GDP (Millions USD) against the Countries and the generated graph is a map. It shows the size of the bubbles with the volume of median GDP.

The median GDP is used to get the distribution of the GDP across 25 years.

In the above visualization, we observe that the United States has the biggest GDP followed by China, Japan, Germany, and United Kingdom. The purpose of this visualization is to focus GDP relation with population and other parameters for these countries.

**Visualization 2**:

Chart, scatter chart

Description automatically generated

The above animation is created by plotting the fertility rate against the life expectancy rate with a population as the size of the bubbles that present countries.

The purpose of this animation is to visualise the overall trend of fertility rate and life expectancy rate with respect to population size.

From the animation, we can observe that the USA and most of the European countries, and Japan which have higher life expectancy rates and lower fertility rate does not change much in the last 25 years. But most of the developing countries including China have moved north significantly with improved life expectancy rates and decreased fertility rates.

It gives some form of impression that with an increased life expectancy the fertility rate decreases.

**Visualization 3**:

Chart, bubble chart

Description automatically generated

This animation is created by plotting population against GDP with a population as the size of the bubbles which represent countries.

The purpose of the animation is to observe the overall trend of GDP with respect to population size.

In the above animation, we can observe that the USA has increased its GDP significantly without increasing its population.

China and India have added population a lot, but China has increased its GDP much more than India.

All the other countries have not changed much. They have improved their GDP without increasing the population by the same proportions.

This clearly indicates that population does not impact the GDP significantly. There can be some other parameters such as technology, land size, education or Govt/foreign policy which decide the GDP of a country.

**Visualization 4**:

Chart, bar chart

Description automatically generated

This visualization is created by plotting the fertility rate and life expectancy rate against a hierarchy. Using the hierarchy, we can drill down from region to country level if we want to focus on any particular region only. The result is a column bar chart.

Here also I have considered the median fertility rate and median life expectancy rate to get the best picture of distribution across 25 years.

In this visualization, we observe that the life expectancy rate is highest in North America, but the fertility rate is lower compared to other regions. The same goes for Europe and Central Asia. The Sub-Saharan Africa region has the highest fertility rate it also has the lowest life expectancy rate.

Let’s focus more on North America and Sub-Saharan Africa countries and analyze the trend.

By drilling down the Sub-Saharan Africa region, we observe 3 interesting data. As expected, Mauritius and Seychelles which have the highest life expectancy rates (Approx. 73 years), have lower fertility rates (1.63, 2.25).

But South Africa is an exception here. Though it has a lower fertility rate of 2.60, the life expectancy rate is very low at approx. 58 years. We can analyze population and GDP data to get more insights.

By drilling down North America, the trend is as expected. The fertility rate and life expectancy rate go hand in hand as shown below.

Chart, bar chart

Description automatically generated

When we drill down the South Asia region where all the countries are developing countries, we see a similar trend.

Chart, bar chart

Description automatically generated

We also see a similar trend in East Asia & Pacific.

We observe few exceptions in Europe & Central Asia region. Faeroe Islands has a life expectancy of 80 years but a fertility rate of 2.50. Iceland has a life expectancy of 82 years but a fertility rate of 2.01. Tajikistan and Uzbekistan have a higher fertility rates of 3.68 and 2.48 compared to other countries with similar life expectancy rates of 68 and 69 years.

In the Middle East & North Africa region, Israel is the exception with a higher life expectancy rate of 81 years but a comparatively higher fertility rate of 2.96.

A summary of all the identified countries is as follows.

A screenshot of a computer

Description automatically generated with low confidence

**In this visualization, we can see many exceptions and there is a clear indication that for a few countries there are other parameters that decide the life expectancy and fertility rate.**

**Visualization 5**:

Chart, bar chart

Description automatically generated

This graph is created by plotting GDP against the population using the region and country hierarchy. The result is a column bar chart.

Here I have used median GDP and maximum population to get the distribution over the 25 years.

In this visualization, we observe that North America has the highest median GDP and median population. But we know population data here does not make much sense as we are not considering the land mass or the number of countries. This parameter can be added for further research.

Let’s focus on all the identified countries in the previous visualization.

Chart, bar chart

Description automatically generated

When we compare this summary chart of selected countries with the summary chart of the previous one, we can observe that though China has the highest population its fertility rate is lower than a similar country like India. It has also gained GDP but much more than India. China’s life expectancy rate is also better than India’s. So, we can assume that China’s GDP is not driven by population but maybe because of other parameters such as export, manufacturing, Govt policy or technology.

India’s GDP seems much dependent on population and not on technology or export which can be confirmed by looking into those data.

The USA and other European countries also do not depend on population for GDP.

However, for all the countries GDP seems impacts life expectancy and hence fertility rate.

Looking at Israel it seems there is some other factor that impacts its higher fertility rate even if it has fewer people and high GDP. This can be researched further.

To validate our findings so far, lets focus on top 5 and bottom 5 countries and analyse the trends.

**Visualization 6**:

Chart, line chart

Description automatically generated

This visualization is created by plotting population, fertility rate, life expectancy rate, and GDP rate across the 25years and including the top 5 countries by population. The result isa line chart.

The purpose of the graph is to observe the trends in the top 5 countries in one place and validate the findings so far.

In this graph, we observe that the populations of China and India have increased a lot during the last 25 years. The GDP of China and the USA has increased more than other countries. The life expectancy of all countries has improved. But when we observe the fertility rate, India and Brazil have a much higher decreased fertility rate than the other 3 countries. This again validates that GDP does not depend on population and fertility rate does not depend on life expectancy rate for all countries.

**Visualization 7**:

Chart, line chart

Description automatically generated

This visualization is created by plotting the population, fertility rate, life expectancy rate, and GDP across the last 25 years. We included the bottom 5 countries only by population size. The result is a line chart.

The purpose of this chart is to observe the trends of the 5 lowest populous countries in one place to validate our findings so far.

In this chart, we observe that the populations of Algeria and Afghanistan have increased the most. The life expectancy rate has improved for all as observed for other countries before. The GDP for Afghanistan has not increased much even if the population has increased but for Algeria, the GDP has increased a lot. There can be some other reason for this that needs to be identified. The fertility rate has decreased for all except Algeria. This is an exception and should be considered separately.

Overall, the findings before still hold.

**Visualization 8**: (**Dashboard I**)

Chart

Description automatically generated with medium confidence

This is the dashboard created with two animations and one map.

The purpose of this dashboard is to observe the world trends in population, GDP, fertility rate and life expectancy rate in one place and get insights in most effective way.

**Visualization 9**: (**Dashboard II**)

Chart, bar chart

Description automatically generated

This dashboard is created with the region and country-level charts.

The purpose of this dashboard is to give flexibility to users to analyze region-level trends which can be drilled down to the country level. Also, there are filters to select a specific region or country to further look into.

**Visualization 10**: (**Dashboard III**)

Chart

Description automatically generated

This dashboard is created with top N and bottom N charts.

The purpose of this dashboard is to give flexibility to the users to verify Top N and Bottom N countries by population. They can select maximum top 20 or bottom 20 countries.

**Conclusion**

I have analyzed the trends of population, GDP, fertility rate, and life expectancy rate of all the countries at the world, region, and country levels. The concluding findings are briefed below.

* The population does not impact GDP overall.
* The life expectancy rate has a significant relationship with the fertility rate. With increased life expectancy the fertility rate comes down.
* Though the population has no impact on GDP, GDP plays a major role on life expectancy rate and hence fertility rate.
* We observed few exceptions for countries such as Israel, India and Algeria. This indicates that there can be some other parameter that plays a role for a few countries.

**Additional Research Questions**

* Is there any other factor such as export, manufacturing, education, or Govt policy that impacts the GDP of any country?
* Is the land size and population per square kilometer (population density) play a role in GDP and other parameters?

**My References**

https://www.data.worldbank.org/