

Connecting the Dots: GDP, Internet Access, Life Expectancy and Population

Group 5 – Svetlana, Ilknur, Karla and David

Dataset

The dataset summarizes 214 countries' region code, country code, GDP, population, Population compound growth rate (CGR), internet users (per 100), percent of population in the largest city, 2014 life expectancy from birth, female literacy rate and exportation rate. Outside of country classifications, there were subsets that included various countries or similar cultures based on the categories mentioned above.

	Country Name	Region Code	Country Code	GDP, PPP (current international \$)	Population, total	Population CGR 1960-2015	Internet users (per 100 people)	Popltn Largest City % of Urban Pop	2014 Life expectancy at birth, total (years)	Literacy rate, adult female (% of females ages 15 and above)	Exports of goods and services (% of GDP)
0	Aruba	MA	ABW	NaN	103,889	1.19%	88.70	NaN	75.50	97.51	NaN
1	Andorra	EU	AND	NaN	70,473	3.06%	96.90	NaN	NaN	NaN	NaN
2	Afghanistan	ME	AFG	62,912,669,167	32,526,562	2.36%	8.30	53.40%	60.40	23.87	0.07
3	Angola	AF	AGO	184,437,662,368	25,021,974	2.87%	12.40	50.00%	52.30	60.74	0.37
4	Albania	EU	ALB	32,663,238,936	2,889,167	1.07%	63.30	27.30%	77.80	96.77	0.27
5	Arab World	NaN	ARB	6,435,291,560,152	392,022,276	2.66%	39.50	29.80%	70.60	NaN	NaN
6	United Arab Emirates	ME	ARE	643,166,288,737	9,156,963	8.71%	91.20	30.80%	77.40	95.08	NaN

Snapshot of raw data set

Objective

From this dataset, the objective was to see whether there was a correlation between GDP and internet penetration, life expectancy and population size. The analysis factors all reported countries, however, the information was analyzed from a regional standpoint.

Target Audience

The target audience for this analysis would be students, researchers, development initiatives, and institutions interested in global development or socio-economic trends. This analysis would prove useful in talks about global development and the role of socio-economic factors in spurring GDP and overall economic growth.

Limitations

The limitations are:

- The dataset provides a snapshot of the countries in 2015 rather than an analysis over several years. Had the analysis covered a wider scope of years there would be greater potential to identify trends or better identify the impact of each factor impact over time.
- Whilst the analysis details the impact of internet penetration, life expectancy and population size, other socio-economic factors play impactful roles in the development of an economy i.e. political stability, infrastructure, education and investment, overall technological advancement, among others.
- Missing values resulted in some columns being deleted. The most notable columns that would have been important to also explore for their effects on GDP were women's literacy and exportation rates. These, among other, missing values restricted a more fulsome analysis (58 countries did not provide women's literacy rates and 69 countries did not provide information on exportation rates).

Data Cleaning

Data cleaning focuses on formatting and ensuring data consistency to properly represent the various regions and their GDP. This involved:

- File identification
Identifying the file type (chardet) that would best help us to read the source csv file i.e. "latin-1"
- Deletion
Removing the additional subsets that would inflate the information i.e. remove data without "Region Codes" e.g. Arab World, middle income, etc.
- Renaming
Renaming and mapping columns to data to have clearer categories
Renaming "NA" region code to "North America" as the region code as it was being read as NaN
- Formatting
Identifying and converting the columns that did not have the correct data types i.e. object to float64
Resetting the index given the deletion of certain rows
Removing characters such as commas, spaces, percent symbol, etc.
Adding commas and decimal to float, in addition to, removing scientific notation for numbers

	Country Name	Region Code	Country Code	GDP (\$)	Population Total	Population CGR 1960-2015	Internet Users per 100	Pop. Largest City (%)	Life Expectancy	Female Literacy Rate (%)	Exportation (% of GDP)
0	Aruba	MA	ABW	NaN	103,889	1.19%	88.70	NaN	75.50	97.51	NaN
1	Andorra	EU	AND	NaN	70,473	3.06%	96.90	NaN	NaN	NaN	NaN
2	Afghanistan	ME	AFG	62,912,669,167	32,526,562	2.36%	8.30	53.40%	60.40	23.87	0.07
3	Angola	AF	AGO	184,437,662,368	25,021,974	2.87%	12.40	50.00%	52.30	60.74	0.37
4	Albania	EU	ALB	32,663,238,936	2,889,167	1.07%	63.30	27.30%	77.80	96.77	0.27
...
207	Samoa	PA	WSM	1,144,453,766	193,228	1.05%	25.40	NaN	73.50	99.12	0.27
208	Yemen, Rep.	ME	YEM	NaN	26,832,215	3.04%	25.10	31.90%	63.80	54.85	NaN
209	South Africa	AF	ZAF	723,515,991,686	54,956,920	2.11%	51.90	26.40%	57.20	93.43	0.31
210	Congo, Dem. Rep.	AF	COD	60,482,256,092	77,266,814	2.99%	3.80	35.30%	58.70	65.90	0.29
211	Zambia	AF	ZMB	62,458,409,612	16,211,767	3.08%	21.00	32.90%	60.00	80.57	NaN

212 rows x 11 columns

Analysis

Finding the mean of GDP, internet penetration, life expectancy, population size. From the results, the following table was summarized.

```
#Summary DataFrame for Regional GDP - Bar chart

summary_region_code = pd.DataFrame({"GDP ($)":region_code_gdp,
                                     "Population Total":region_code_population,
                                     "Life Expectancy":region_code_life,
                                     "Internet Users per 100":region_code_internet
                                     })

summary_region_code.head(13)
```

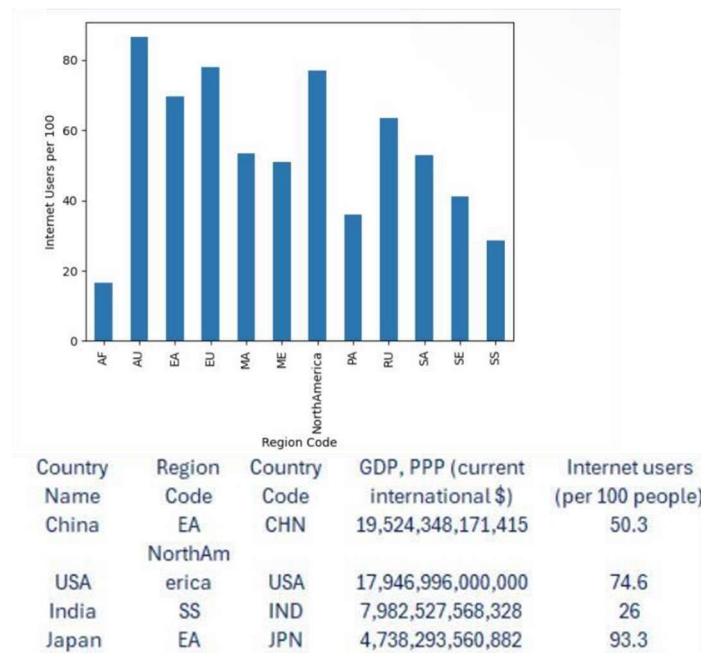
	GDP (\$)	Population Total	Life Expectancy	Internet Users per 100
Region Code				
AF	93,127,962,064.26	21,178,547.15	59.63	16.46
AU	626,169,928,411.00	14,188,434.50	81.85	86.40
EA	4,421,207,403,156.17	226,400,467.71	77.97	69.57
EU	532,701,497,414.36	12,776,690.60	78.66	77.91
MA	143,511,735,523.35	7,159,446.40	74.90	53.29
ME	375,566,320,177.30	23,176,709.15	72.80	50.93
NorthAmerica	9,767,796,223,481.50	119,108,902.67	80.45	76.90
PA	2,225,181,311.60	668,781.19	71.43	35.85
RU	952,994,953,561.00	40,111,218.25	72.65	63.45
SA	539,671,120,224.36	34,847,973.92	73.88	52.96
SE	665,659,997,551.20	57,430,696.09	71.98	41.05
SS	1,399,890,889,534.14	244,519,248.00	70.93	28.61

Findings

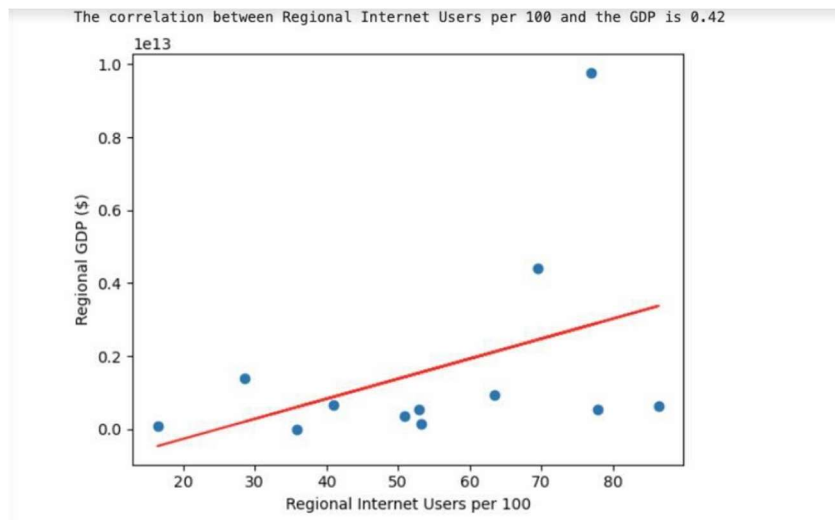
GDP vs Internet

In this Project we have used the dataset where internet use was captured per 100 users count, thus giving us more of a percentile data, than actual numbers.

Though many people have extolled the virtues of internet usage in reducing poverty and improving the quality of life, we wanted to see using our Data set whether GDP had a strong correlation with Internet usage of each region or not. Our Dataset indicates 13 Regions with countries having variable spreads in GDP. As well, we realized that although some individual countries are leaders in GDP, their Internet Use per 100 will give us some food for thought.



Taken those regions means, we have created a chart indicating correlation between GDP of the Region to the number of Internet users. We have received a moderate positive correlation of 0.42, which is telling us that while the correlation is positive, in the year of 2015 it was certainly not very strong. Considering more recent data sets of the similar exposure were not available, yet number of Internet Users per 100 in each region has inevitably grown, we think that for countries to invest in the increasing of Internet usage will be beneficial not only from the education point of view, but also for local businesses having a better chance to grow, by which increasing their regions' GDP after all. At last, we noticed that while one of the biggest challenges with this data set was lack of census data of countries participating, streamlining the process of data gathering would certainly increase transparency in such analysis.



GDP vs Life Expectancy

The bar chart illustrates average life expectancy by region. Regions like AU (Australia) and EU (Europe) exhibit the highest life expectancy, likely due to advanced healthcare systems, higher living standards, and stable economies.

Life expectancy in North America aligns with the highest-performing regions, likely due to advanced healthcare and socioeconomic stability.

Regions like ME (Middle East) and PA (Pacific Asia) show moderate life expectancy, indicating a balance between economic development and access to healthcare. These regions are improving but still lag high-income regions.

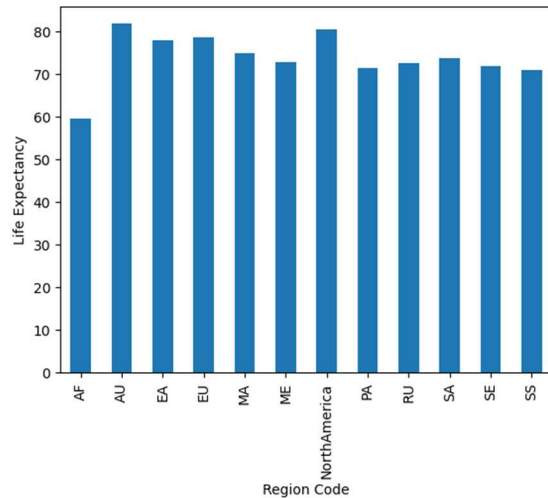
Regions like AF (Africa) and SS (Sub-Saharan Africa) have significantly lower life expectancy, reflecting challenges such as poverty, limited healthcare access, and higher prevalence of diseases.

Development Disparities:

Life expectancy is a strong indicator of regional development, with wealthier regions offering better infrastructure for health and well-being.

Policy Focus:

Lower-performing regions like AF and SS would benefit from targeted investments in healthcare, education, and disease prevention to improve longevity.

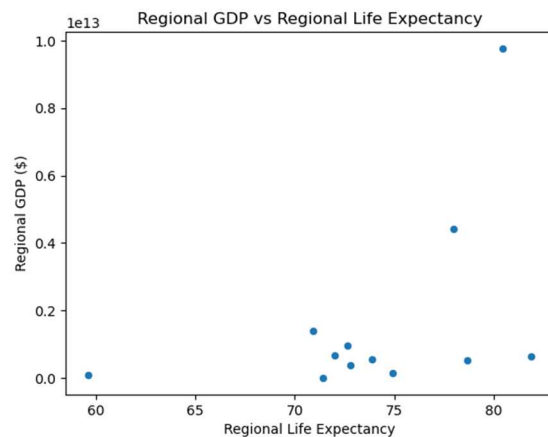


Scatter GDP vs Life Expectancy Plot

Most countries with lower life expectancy (below 70 years) have relatively low GDP values.

There are clusters of countries with life expectancy between 70 and 85 years where GDP is concentrated at moderate levels, indicating that higher life expectancy generally corresponds to stronger economies, though not universally.

While there may be a general positive association between life expectancy and GDP, it is not linear or strong, suggesting the influence of additional factors like population size, industrialization, and resource wealth.



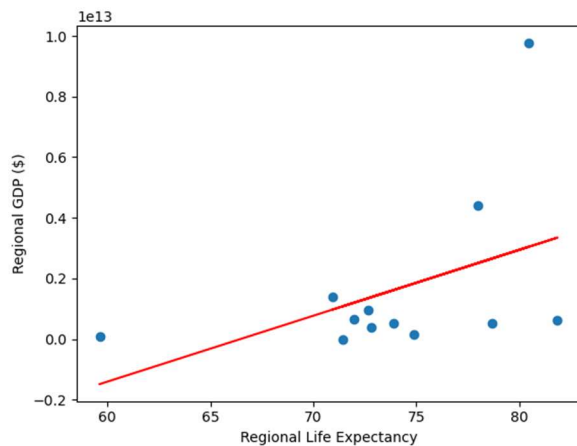
Regional GDP vs Regional Life Expectancy Regression (Correlation) Chart:

As life expectancy increases, GDP tends to rise, although the correlation is not very strong, indicating other factors also significantly impact GDP.

There are potential outliers with high GDP but slightly below-average life expectancy. These could be countries with strong economies due to factors like resource wealth or technological dominance, despite lower life expectancy.

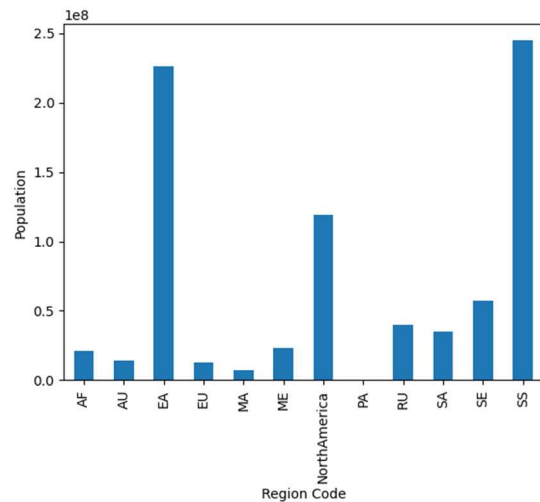
The moderate correlation suggests life expectancy alone does not dictate GDP. Other elements, such as infrastructure, education, political stability, and natural resources, play crucial roles.

The weak positive correlation observed between life expectancy and GDP in the analysis reflects that higher life expectancy supports economic productivity. However, the relationship also depends on other factors like education, healthcare quality, and economic policies.

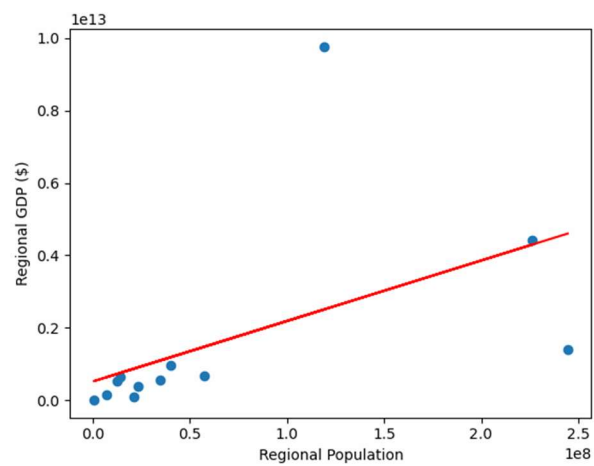


GDP vs Population size

GDP vs. Population bar graph shows the important relationship. Demographic factors can significantly influence economic productivity and growth. From these charts we can see that the top 3 regions are EA, SS and North America. These nations effectively utilize their large populations to increase the labour supply and market size, driving economic growth. This can boost a country's GDP through increased consumption and workforce participation.



The positive correlation shown in the scatter plot shows that this can lead to more access to labor, which can increase productivity and output. More people in the country will also lead to more demand for goods. Demanding and producing more goods will yield economic growth. This is an example of population growth having a positive effect on the economy.



Conclusion

There were moderately positive correlations between GDP and the individual factors of internet usage, life expectancy and population size. All factors could boost the GDP of a country, but they would not be able to do so independently. They would all rely on other socio-economic factors being present. A nation is built on several moving parts and other socio-economic factors such as infrastructure, healthcare, government investment, political stability, etc.

For future research, widening the scope to cover several years and including other socio-economic factors would provide a greater insight into the effects of GDP growth over time. However, this snapshot provides an impactful preview at how GDP can be boosted by having one or more factors that are favorable for expansion.