Introduction to Data Science AI

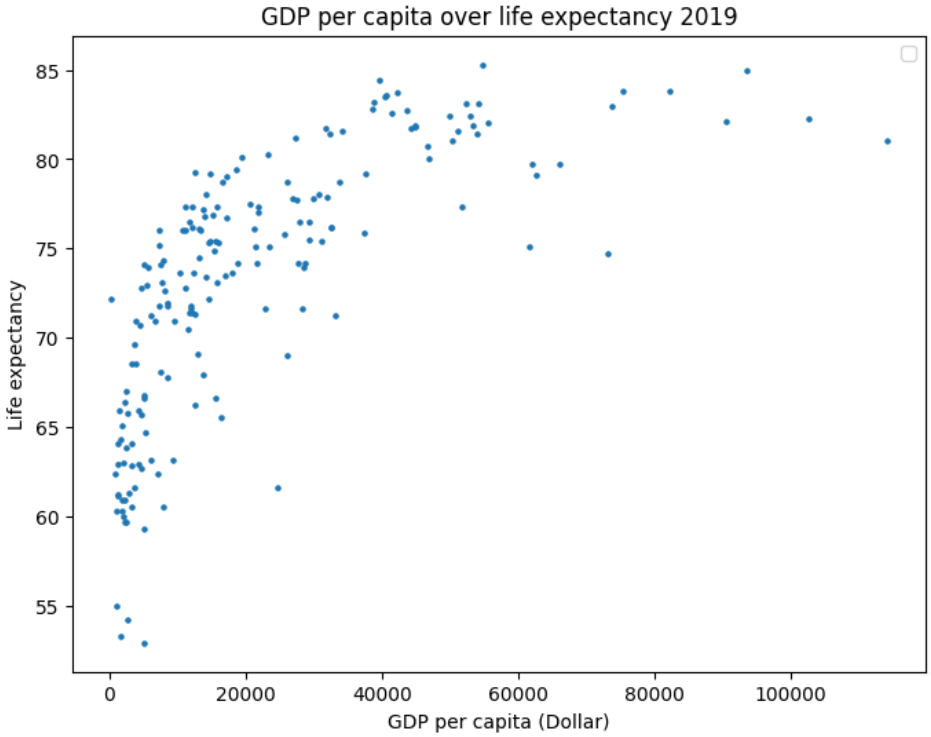
Johanna Norell

March 1, 2024

1 Data Wrangling

The data that was imported had to be wrangled in order for it to be used effectively. Firstly the data was wrangled to only include GDP per Capita and Life Expectancy from 2019 which was the latest year the data was available. Some columns where renamed in order to better reflect their content. The entities which represented regions other than countries were also removed. Lastly, the two separate DataFrames where merged too one bigger DataFrame that included both Life Expectancy and GDP per Capita.

Here is the scatter plot that was constructed using Python which shows the GDP per capita in relation to life expectancy in the world.



Data Interpretation

1. Which countries have a life expectancy higher than one standard deviation above the mean?

Countries
Australia
Austria
Belgium
Bermuda
Canada
Cyprus
Denmark
Finland
France
Germany
Greece
Hong Kong
Iceland
Ireland
Israel
Italy
Japan
Luxembourg
Macao
Malta
Netherlands
New Zealand
Norway
Portugal
Qatar
Singapore
Slovenia
South Korea
Spain
Sweden
Switzerland
Taiwan
United Kingdom

Table 1: List of Countries with a life expectancy higher than one standard deviation above the mean

The list comprises 33 countries, each with a life expectancy greater than one standard deviation above the mean. These countries account for 18% of the total number of countries in the dataset, which consists of 183 countries. Under a normal distribution, 16% would be expected to lie more than one standard deviation above the mean. This suggests that our data is slightly skewed to the left.

1.1 Interpreting variable *life expectancy*

2. Which countries have high life expectancy but have low GDP (note the difference between GDP and GDP per capita)? Motivate how you have chosen to define "high" and "low."

For this question GDP data had to be used. This was taken from https://ourworldindata.org/economic-growth. It also had to be wrangled a bit to only contain the year 2019.

Since the difference in GDP vary so much, the median was used instead of mean. We chose to define **low GDP** as below the median divided by two. Additionally, we defined **high life expectancy** as one standard deviation above the mean. The results, together with GPD per capita, can be seen in the table above.

Table 2: 'Low' GDP and 'High' Life Expectancy (2019)

Entity	Code	Year	GDP (MUSD)	GDP per Capita (USD)	Life Expectancy
Bermuda	BMU	2019	3142.3936	50273.473	81.0
Cyprus	CYP	2019	28053.9690	32301.820	81.4
Iceland	ISL	2019	17972.6250	53011.742	82.4
Malta	MLT	2019	17134.9080	38910.080	83.2

The result shows us that four countries have a low GDP but high life expectancy, according to the chosen definitions. However, one should notice that all countries have quite high GDP per Capita - in fact all countries above the mean of 22,032 USD/capita.This suggest that the low GDP relates to a low number of inhabitants in the countries rather than a poor population. Followingly, it is not as suprising that these countries have high life expectancy since they are not poor in terms of GDP per Capita.

1.2 Interprating variable *economy*

3. Does every strong economy (normally indicated by GDP) have high life expectancy?

The mean life expectancy is 73.4 and as can be seen in the table there are two countries with a lower life expectancy. Thus it can be said somewhat confidently that a high GDP leads to a higher life expectancy.

Table 3: Top 10 Countries by GDP (2019)

Entity	Code	Year	GDP_PC (MUSD)	Life Expectancy	GDP (USD)
United States	USA	2019	62589.000	79.1	20595845.0
China	CHN	2019	14128.812	78.0	20257660.0
India	IND	2019	6711.385	70.9	9170555.0
Japan	JPN	2019	39704.234	84.4	5036891.0
Germany	DEU	2019	51190.890	81.6	4275312.0
Russia	RUS	2019	28526.293	73.9	4161194.4
Indonesia	IDN	2019	11595.102	70.5	3137931.0
Brazil	BRA	2019	14593.961	75.3	3080048.6
United Kingdom	GBR	2019	44274.957	81.7	2989895.5
France	FRA	2019	43755.062	82.7	2946958.4

1.3 Indicators

4. Related to the above question (question 3), what happens if you use GDP per capita as an indicator of a strong economy as opposed to GDP alone? Explain the results you obtain through this analysis, and discuss any insights you get from comparing these results to question 3.

Defining **Low** GDP per Capita as below the mean and **High** life expectancy as one standard deviation divided by 1.5 above the mean. This was chosen because there was no countries with one standard deviation above the mean.

Table 4: 'Low' GDP per Capita and 'High' Life Expectancy (2019)

Entity	Code	Year	GDP (MUSD)	GDP per Capita (USD)	Life Expectancy
Albania	ALB	2019	36103.04	12531.788	79.3
Antigua and Barbuda	ATG	2019	1603.8545	16514.492	78.7
Costa Rica	CRI	2019	93491.83	18522.178	79.4
Lebanon	LBN	2019	101120.85	14749.867	79.2
Maldives	MDV	2019	10306.475	19411.275	80.1
Thailand	THA	2019	1191732.8	17116.309	79.0

When using GDP per Capita as an indicator for strong economy, we can see that countries with less strong economy have slightly lower life expectancy than the countries having less strong GDP. This indicates that GDP per Capita might be a better way of representing the strength of the economy. Though, table 4 tells us that even though these countries have a low GDP per Capita, according to our definition, their life expectancy is above average (which is 73.4 years as mentioned above). This indicates it might not be as significant that a less strong economy relates to a lower life expectancy, in relation to the opposite relation.

Table 5: Top 10 highest GDP per Capita (2019)

Entity	Code	Year	GDP (MUSD)	GDP per Capita (USD)	Life Expectancy
Qatar	QAT	2019	323141.14	114100.810	81.0
Ireland	IRL	2019	501053.6	102622.450	82.3
Macao	MAC	2019	59874.165	93488.375	85.0
Luxembourg	LUX	2019	55710.794	90479.410	82.1
Singapore	SGP	2019	477907.88	82336.340	83.8
Switzerland	CHE	2019	646919.6	75298.820	83.8
Norway	NOR	2019	396253.9	73668.790	83.0
Brunei	BRN	2019	31737.774	73249.190	74.7
United Arab Emirates	ARE	2019	645956.24	66112.720	79.7
United States	USA	2019	20595845	62589.000	79.1

In table 5, the relation between a high GDP per Capita and a high life expectancy is clear. Every country in the top 10 has a higher life expectancy than the mean. This tells us that a high GDP Per Capita represents a countries economy better than only the GDP. A small but rich country will have a lower GDP than a large but poor country. The Life Expectancy is thus higher in the top 10 for GDP per Capita than only GDP.

References

[1] Roser, M., Arriagada, P., Hasell, J., Ritchie, H., & Ortiz-Ospina, E. (2019). Economic Growth. Retrieved August 26, 2023, from https://ourworldindata.org/economic-growth

[2] Roser, M., Ortiz-Ospina, E., & Ritchie, H. (2019). Life Expectancy. Retrieved August 26 2023, from https://ourworldindata.org/life-expectancy DAT565 Introduction to Data Science and AI (LP1 2023-2024) Page 2