

Finding Happiness

An Analysis of World Happiness

By: Erik Luu

The purpose of life is to survive and pass on our genetics, but that's boring. Is the purpose of life to be happy? What is happiness? What is the root of happiness?

One of the most well-known phrases in the United States Declaration of Independence is "Life, Liberty, and the pursuit of Happiness." In essence, "no person or class of persons shall be denied the same protection of the laws that is enjoyed by other persons or other classes." Happiness must be important if it's one of the three things that can't be taken away from us.

There are two ways to analyze happiness: by the individual or by the community. The World Happiness Report is a publication of the Sustainable Development Solutions Network powered from the *Gallup World Poll* and *Lloyd's Register Foundation*. It surveys the conditions of happiness by country and ranks them by how a sample of citizens subjectively measure their welfare. From the years 2015-2019, there was a typical annual sample for each country of about 1,000-3000 people from 156-159 of the world's countries. The data comes from the Gallup World Poll surveys. It asks respondents to think on scale from 0 to 10, where 10 is the best possible life and 0 is the worst.

The rest of the features analyzed in this report can be explained as follows:

- **GDP per Capita¹**: The Gross Domestic Product divided by the midyear population. Essentially a country's economic output divided by its population.
- **Social Score²**: Basic Human Needs, Foundation of Wellbeing, and Opportunity.
- **Life Expectancy³**: The overall mortality level of a population.
- **Official Languages, Population, Region⁴**: Self-explanatory.

Data Collection and Cleaning

I retrieved my initial data from a Kaggle dataset of the World Happiness Report⁵. Formatting this data was simple as the provided years, 2015-2019, were all in csv files. I merged them together and grouped by year to simplify the code and only needed to fill a few missing data points manually.

¹ <https://www.imf.org/>

² https://en.wikipedia.org/wiki/Social_Progress_Index

³ <https://worldpopulationreview.com/countries/life-expectancy>

⁴ <https://restcountries.com/>

⁵ <https://worldhappiness.report/>

	Overall rank	Country	Score	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption	Year
0	1	Finland	7.769	1.34000	1.58700	0.98600	0.59600	0.15300	0.39300	2019
1	2	Denmark	7.600	1.38300	1.57300	0.99600	0.59200	0.25200	0.41000	2019
2	3	Norway	7.554	1.48800	1.58200	1.02800	0.60300	0.27100	0.34100	2019
3	4	Iceland	7.494	1.38000	1.62400	1.02600	0.59100	0.35400	0.11800	2019
4	5	Netherlands	7.488	1.39600	1.52200	0.99900	0.55700	0.32200	0.29800	2019

Figure 1

Shown above in Figure 1 is the head of the data frame. The Happiness Score, labeled “Score”, comes from the Gallup World Poll surveys. “... it asks respondents to think of a ladder, with the best possible life for them being a 10, and the worst possible life being a 0.” While the sample size seems small, it is large enough to give a good estimate at the national level. This can be confirmed by the 95% confidence intervals from their data (not shown above). The numbers for each column, besides “Score”, are coefficients for weighting each attribute. For example, the variable of life expectancy in the case of Finland in the year 2021 is 0.741. This number is obtained by calculating the difference of healthy life expectancy in Finland and the country with the lowest life expectancy and multiply by the given coefficient (fig. 2). I did not find this information too intuitive but it useful for comparison with my analysis. The data is still useful to visualize correlations and linear regression lines of each attribute with the Happiness Score to see which attributes are best to predict happiness and if they will be similar to the World Happiness Report’s weights.



Figure 2

Toying with the project in my mind, I had other ideas of factors outside of the main GDP, social support, and life expectancy that could possibly show some interesting results. I used an additional API called *RestCountries*⁶, of which I obtained the official languages, population, latitude and longitude, independence, landlocked information, cca3, and region for each country; filling in the missing values as I went.

The last bit of data collection was a lot of web scraping. I retrieved the GDP in billions for each country, three scores of “Basic Human Needs”, “Foundation of Wellbeing”, and “Opportunity”, which average to the overall “Social Score”, and life expectancy (fig. 3). The social scores come from the *Social Progress Index* which measures the extent to which countries provide for the social and environmental needs of their citizens. However, this index has been criticized for using data biased against the global

⁶ <https://restcountries.com/>

south and basing too much on European values. For example, basic human needs can be extremely different in countries across the world.

	Overall rank	Score	Country	GDP (Billions)	Social Score	Basic Human Needs	Foundation of Wellbeing	Opportunity	Life Expectancy	language
0	1	7.769	Finland	296.016	91.89	96.22	91.29	88.15	82.312	{'fin': 'Finnish', 'swe': 'Swedish'}
1	2	7.600	Denmark	396.666	92.11	96.11	91.58	88.66	81.256	{'dan': 'Danish'}

Figure 3

Data Exploration

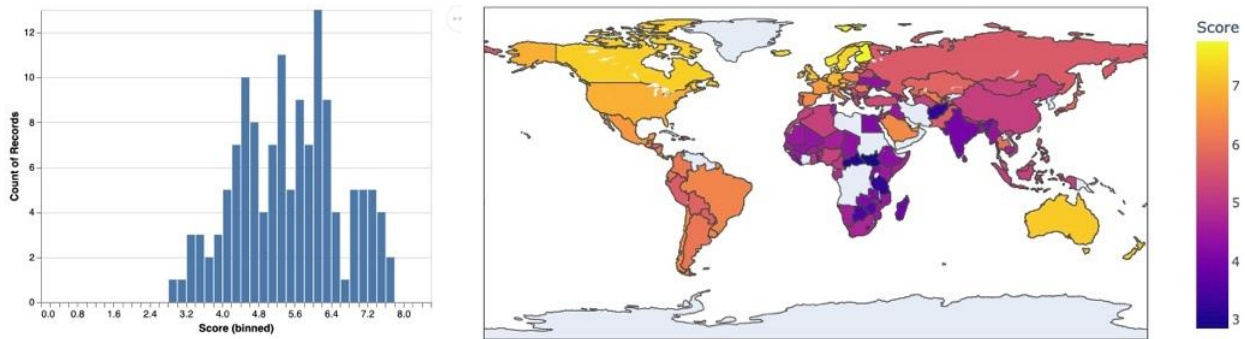


Figure 4 Distribution of Happiness Scores (Left) Map of Happiness Scores (Right)

From a glance, the west is considerably happier. In figure 5 below. The countries with higher scores are those with a higher GDP per capita. Regions such as Europe, North America, and Australia all have the highest GDPs in the world and have some of the happiest scores. Coincidence? Money opens doors to education, health care, and time. Life expectancy matches the same trend as seen in figure 6.



Figure 5. Scatterplot of GDP vs Happiness (Left) Map of GDP per Capita (Right)



Figure 6. Life Expectancy vs Happiness (Left) Map of Life Expectancy (Right)

I am currently in this class trying to get the best grade I can to raise my GPA to eventually graduate to show off to employers, all to get a job (aka Money). However, there are plenty of countries that have high foundations of wellbeing and life expectancies without an incredibly large GDP per capita. There is a huge logarithmic slow after a certain GDP per capita threshold. Western Europe sticks out as on average it may have a larger GDP per capita, but Central and Eastern Europe scores roughly the same wellbeing and life expectancy scores with a lower average GDP.

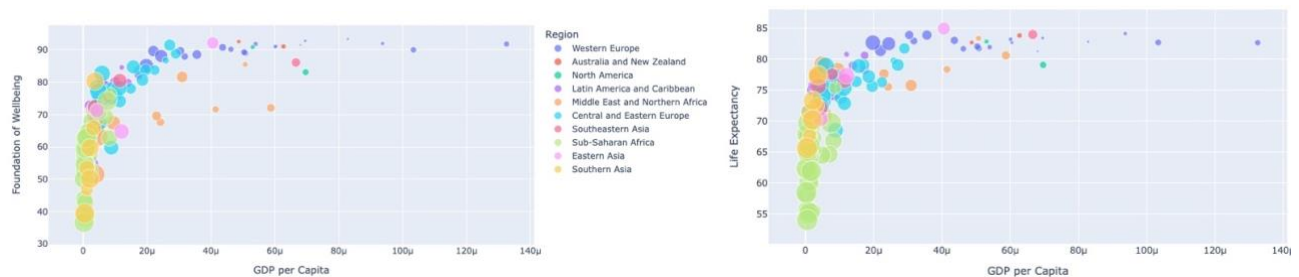


Figure 7. GDP vs Foundation of Wellbeing (Left) GDP vs Life Expectancy (Right)

While all variables positively correlate with GDP per capita, social score correlates most with the end happiness score as seen in the correlations grid below and the rest of the variables. Maybe there is hope that there is more to life than just money and I don't need to stress about upholding the stereotype that CS majors are only in it for the money... that would still be nice though.

	Score	GDP per Capita	Social Score	Basic Human Needs	Foundation of Wellbeing	Opportunity	Life Expectancy
Score	1.000000	0.915738	0.966939	0.892055	0.960799	0.957500	0.915779
GDP per Capita	0.915738	1.000000	0.900570	0.783389	0.877464	0.951435	0.818689
Social Score	0.966939	0.900570	1.000000	0.950753	0.988504	0.967837	0.950816
Basic Human Needs	0.892055	0.783389	0.950753	1.000000	0.915191	0.849814	0.980666
Foundation of Wellbeing	0.960799	0.877464	0.988504	0.915191	1.000000	0.960804	0.911354
Opportunity	0.957500	0.951435	0.967837	0.849814	0.960804	1.000000	0.871558
Life Expectancy	0.915779	0.818689	0.950816	0.980666	0.911354	0.871558	1.000000

Figure 8. Average correlations between variables by region.

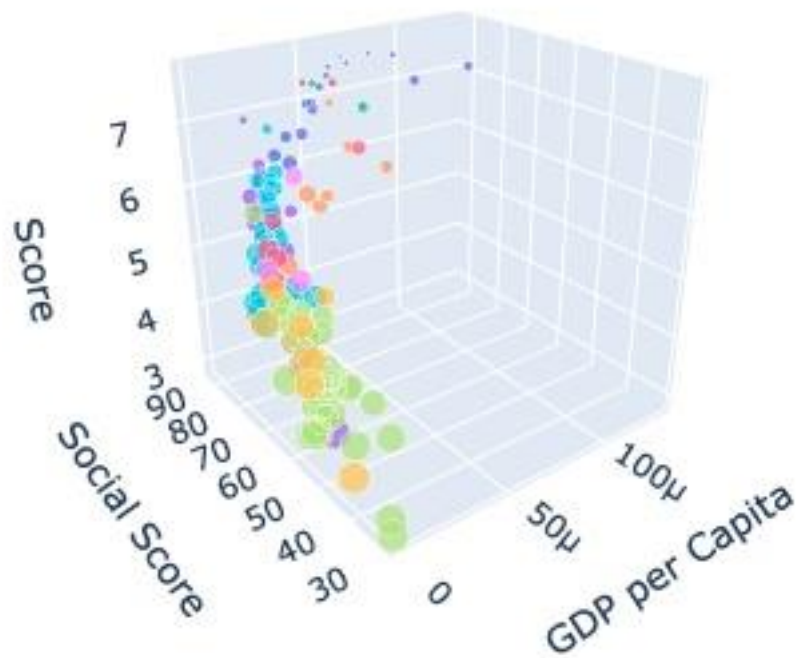


Figure 9. GDP per Capita vs Social Score vs Score

Machine Learning:

I fit two models to this data after splitting the data into 80% training and 20% testing: Linear Regression and KNeighbors Regressor. I chose to use R-Squared to test the metrics of my models primarily because it captures the fraction of variance rather than just the residual error (MSE).

The Linear Regression outperformed KNeighbors on every iteration and combination of variables. The best being GDP per capita, social score, life expectancy, latitude, and region as seen in figure 10. This makes sense as basic human needs, foundation of wellbeing, and opportunity are a part of the social score itself. Latitude also makes sense in the context of history. With high temperatures, plants have adapted to use less water and wilt at higher heats. It is a great place for agriculture and was taken advantage of, along with its people, by more prosperous nations leaving a lasting affect today as slavery, specifically American slavery, was still going strong as of 150 years ago.

What I find interesting from the error values below in figures 10 and 11, is that on its own, using GDP per capita as a predictor compared to the Social Score is worse in both cases. On average, countries with less GDP per capita and a greater social score are happier than countries with a greater GDP per capita and a worse social score. As shown in the 3D plot of figure 9, there are a lot more “happier” countries with lower GDP per capita and high social scores.

['GDP (Billions)', []]	-0.698355
['GDP per Capita', []]	0.393025
['Opportunity', []]	0.468700
['Basic Human Needs', []]	0.562535
['Social Score', 'Basic Human Needs', 'Foundation of Wellbeing', 'Opportunity', []]	0.585650
['Life Expectancy', []]	0.587821
['Foundation of Wellbeing', []]	0.590190
['Social Score', []]	0.604730
['GDP per Capita', 'Social Score', 'Life Expectancy', ['Region']]	0.666587
['GDP per Capita', 'Social Score', 'Life Expectancy', []]	0.667722
['GDP per Capita', 'Social Score', 'Life Expectancy', 'lat', ['Region']]	0.686628

Figure 10. Linear Regression errors for sets of variables.

['GDP (Billions)', []]	0.234347
['Opportunity', []]	0.422894
['Foundation of Wellbeing', []]	0.508452
['Life Expectancy', []]	0.515431
['GDP per Capita', []]	0.541171
['Social Score', []]	0.561265
['Basic Human Needs', []]	0.570699
['GDP per Capita', 'Social Score', 'Life Expectancy', ['Region']]	0.610375
['GDP per Capita', 'Social Score', 'Life Expectancy', 'lat', ['Region']]	0.617992
['Social Score', 'Basic Human Needs', 'Foundation of Wellbeing', 'Opportunity', []]	0.627064
['GDP per Capita', 'Social Score', 'Life Expectancy', []]	0.650324

Figure 11. KNeighbors Regression w/ K=5 (most optimal K) errors for sets of variables.

I decided to use my model to compare countries to themselves with higher and lower inputs. I discovered when a country is happier, GDP affects the predicted happiness scores the most. For example, the United States was ranked number 19. In the situation where everything was the same except the GDP per capita was the 25th percentile, the score dropped a whole point. With the 25th percentile of social score and life expectancy, the score did not drop nearly as much. On the other hand, Afghanistan was ranked 154 out of 156 and using the 75th percentile scores, GDP per capita increased their happiness score from 4.07 to 4.4 while social score increased their happiness score to 5.13 (fig. 12).

[6.93949147]	US – actual: 6.9
[5.94733076]	US 25percentile GDP per Capita
[6.20276051]	US 25percentile Social Score
[6.5574029]	US 25percentile Life Exp
[4.07728149]	Afg – actual: 3.2
[4.39724777]	Afg 75percentile GDP per Capita
[5.13310857]	Afg 75percentile Social Score
[4.32534798]	Afg 75percentile Life Expectancy

Figure 12. Predicted Happiness Scores w/ variable inputs.

Conclusion:

With so many kinds of people, cultures, and values, happiness is a difficult thing to quantify. It can be generalized social support is more impactful for countries who don't have it and GDP per capita is more important for countries who have money. But, then again, everyone has different definitions of happiness, and it takes a combination of the right resources to have a healthy population. It is a balance of necessities and wants. We want social support, we want opportunity, we want welfare, and we want long lives. A combination of GDP per capita and social support is necessary to be happy. With money comes opportunity... literally (fig 12). It is all a big positive feedback loop.

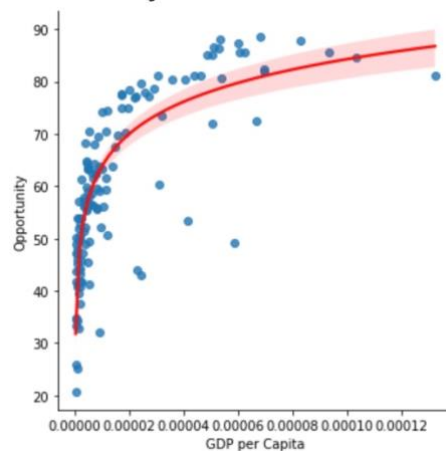


Figure 12. GDP per Capita vs Opportunity

A quote I discovered in my research:

“The purpose of life is not to be happy. It is to be useful, to be honorable, to be compassionate, to have it make some difference that you have lived and lived well.”

-Ralph Waldo Emerson-

Happiness is not a goal, but a byproduct. Life is a balance of work and play, the uncomfortable and the comfortable, the needs and the wants, GDP per capita and social support and life expectancy (and a little latitude). In the end, we are who we are. We do what we need to do to try and satisfy ourselves and find our happiness.