

What makes a Country Happy?

We would like to identify what factors that impact the happiness of a country as measured by the [world happiness report](#) score and whether the historical trends of these factors can predict a country's happiness.

Understanding the factors at a country level can potentially help us identify the happiness level of individuals which can be an important variable in predicting individual behaviour.

The world has experienced a tremendous amount of change in the last 40 years. For example, the economy has transformed significantly with the rise of China and the loss of factory jobs in developed countries. The economy has shifted from a manufacturing economy to a knowledge economy due to what has been dubbed the 4th industrial revolution driven by the internet, technology and data.

The changes the world has experienced has impacted the happiness of countries and we would like to understand what drives happiness. Understanding these factors will help direct policy in the direction that maximizes happiness and better takes care of a societies' people in the face of tremendous disruption.

World Happiness Report

The World Happiness Report is published annually and surveys the population of 149 countries to measure their happiness level and is quantified as a ladder score.

We also combined this data with additional country related metrics from the Gapminder website to create a dataset of features to predict the happiness score. Below is a description of the Ladder score and variables we included in the initial dataset.

Data Description

Ladder score: Happiness score or subjective well-being (variable name ladder): The survey measure of SWB is from the Feb 26, 2021 release of the Gallup World Poll (GWP) covering years from 2005 to 2020. Unless stated otherwise, it is the national average response to the question of life evaluations. The English wording of the question is "Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?" This measure is also referred to as Cantril life ladder, or just life ladder in our analysis.

Logged GDP per capita: The statistics of GDP per capita (variable name gdp) in purchasing power parity (PPP) at constant 2017 international dollar prices are from the October 14, 2020 update of the World Development Indicators (WDI). The GDP figures for Taiwan, Syria, Palestine, Venezuela, Djibouti and Yemen are from the Penn World Table 9.1.

- – GDP per capita in 2020 are not yet available as of December 2020. We extend the GDP-per-capita time series from 2019 to 2020 using country-specific forecasts of real GDP growth in 2020 first from the OECD Economic Outlook No 108 (December 2020) and then, if missing, forecasts from World Bank's Global Economic Prospects (Last Updated: 06/08/2020). The GDP growth forecasts are adjusted for population growth with the subtraction of 2018-19 population growth as the projected 2019-20 growth.

Social support: Social support (or having someone to count on in times of trouble) is the national average of the binary responses (either 0 or 1) to the GWP question "If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?"

Healthy life expectancy: Healthy Life Expectancy (HLE). Healthy life expectancies at birth are based on the data extracted from the World Health Organization's (WHO) Global Health Observatory data repository (Last updated: 2020-09-28). The data at the source are available for the years 2000, 2005, 2010, 2015 and 2016. To match this report's sample period (2005-2020), interpolation and extrapolation are used.

Freedom to make life choices: Freedom to make life choices is the national average of responses to the GWP question "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?"

Generosity: Generosity is the residual of regressing national average of response to the GWP question "Have you donated money to a charity in the past month?" on GDP per capita.

Perceptions of corruption: Corruption Perception: The measure is the national average of the survey responses to two questions in the GWP: "Is corruption widespread throughout the government or not?" and "Is corruption widespread within businesses or not?" The overall perception is just the average of the two 0-or-1 responses. In case the perception of government corruption is missing, we use the perception of business corruption as the overall perception. The corruption perception at the national level is just the average response of the overall perception at the individual level.

gini: description: Gini shows income inequality in a society. A higher number means more inequality.
source_url: <http://gapm.io/ddgini>

income_per_capita: description: Gross domestic product per person adjusted for differences in purchasing power (in international dollars, fixed 2011 prices, PPP based on 2011 ICP).
source: Gapminder based on World Bank, A. Maddison, M. Lindgren, IMF & more.
source_url: <http://gapm.io/dgdppc>

fertility: description: Total fertility rate. The number of children that would be born to each woman with prevailing age-specific fertility rates. source_url: <http://gapm.io/dtfr>

child_mortality: description: Death of children under five years of age per 1,000 live births
source_url: <http://gapm.io/du5mr>

labor_ppt_25-54: description: For age group 25-54, percentage of all labour to total population.
source_url: <https://www.ilo.org/ilostat/>

Exploratory Data Analysis

	Country name	Regional indicator	Ladder score
0	Finland	Western Europe	7.842
1	Denmark	Western Europe	7.620
2	Switzerland	Western Europe	7.571
3	Iceland	Western Europe	7.554
4	Netherlands	Western Europe	7.464

The top 5 happiest countries are from Western Europe, more specifically the scandinavian countries.

	Country name	Regional indicator	Logged GDP per capita
7	Luxembourg	Western Europe	11.647
30	Singapore	Southeast Asia	11.488
14	Ireland	Western Europe	11.342
2	Switzerland	Western Europe	11.117
23	United Arab Emirates	Middle East and North Africa	11.085

Switzerland was the only top 5 country by GDP that is also in the top 5 in happiness score

	Country name	Regional indicator	Social support
3	Iceland	Western Europe	0.983
89	Turkmenistan	Commonwealth of Independent States	0.983
0	Finland	Western Europe	0.954
5	Norway	Western Europe	0.954
1	Denmark	Western Europe	0.954

4 of the top 5 countries in social support also are in the top 5 in happiness ranking. Social support might be an important variable for explaining happiness

	Country name	Regional indicator	Healthy life expectancy
30	Singapore	Southeast Asia	76.953
52	Japan	East Asia	75.100
25	Spain	Western Europe	74.700
2	Switzerland	Western Europe	74.400
20	France	Western Europe	74.000

the top 5 countries by life expectancy are not in the top 5 in happiness score. There was not a large correlation between life expectancy and happiness. Perhaps quality of life means more than length of life.

	Country name	Regional indicator	Freedom to make life choices
38	Uzbekistan	Commonwealth of Independent States	0.970
5	Norway	Western Europe	0.960
105	Cambodia	Southeast Asia	0.959
3	Iceland	Western Europe	0.955
0	Finland	Western Europe	0.949

Finland and Iceland are here as well as in the top 5 happiness score table. Freedom turned out to be important in explaining happiness.

	Country name	Regional indicator	Perceptions of corruption
30	Singapore	Southeast Asia	0.082
137	Rwanda	Sub-Saharan Africa	0.167
1	Denmark	Western Europe	0.179
0	Finland	Western Europe	0.186
6	Sweden	Western Europe	0.237

Lower perceptions of corruption appear to be linked to higher levels of happiness based on top 5 countries for both metrics

	Country name	Regional indicator	income_per_capita
7	Luxembourg	Western Europe	95800.0
30	Singapore	Southeast Asia	91400.0
14	Ireland	Western Europe	75700.0
5	Norway	Western Europe	68200.0
23	United Arab Emirates	Middle East and North Africa	65100.0

The top 5 for income_per_capita is the same as gdp per capita.

	Country name	Regional indicator	gini
101	Ukraine	Commonwealth of Independent States	25.0
27	Slovenia	Central and Eastern Europe	25.4
17	Czech Republic	Central and Eastern Europe	25.9
61	Moldova	Commonwealth of Independent States	26.4
41	Kazakhstan	Commonwealth of Independent States	26.9

more equal countries are not in the top 5 for happiness score

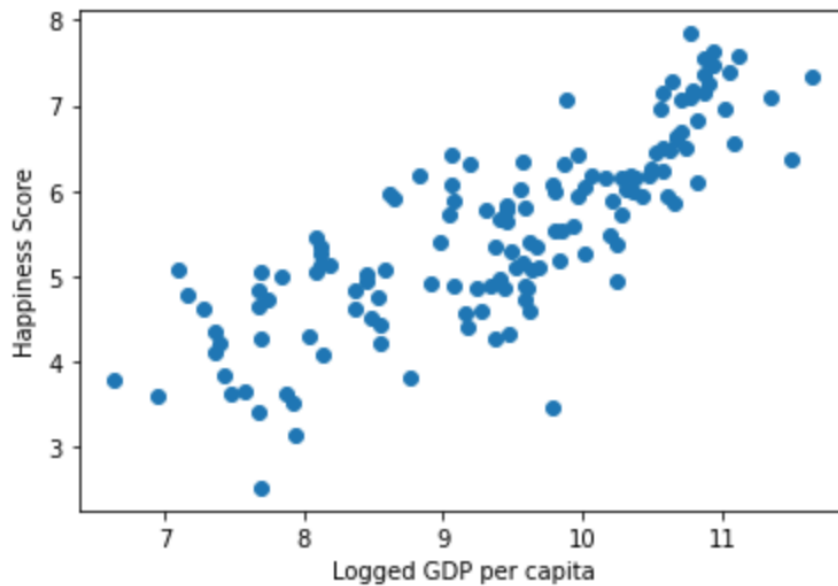
Modeling

We ran a regular linear regression model with all variables, one with only the 5 significant variables, and one with the 5 significant variables as well as the regional indicator columns.

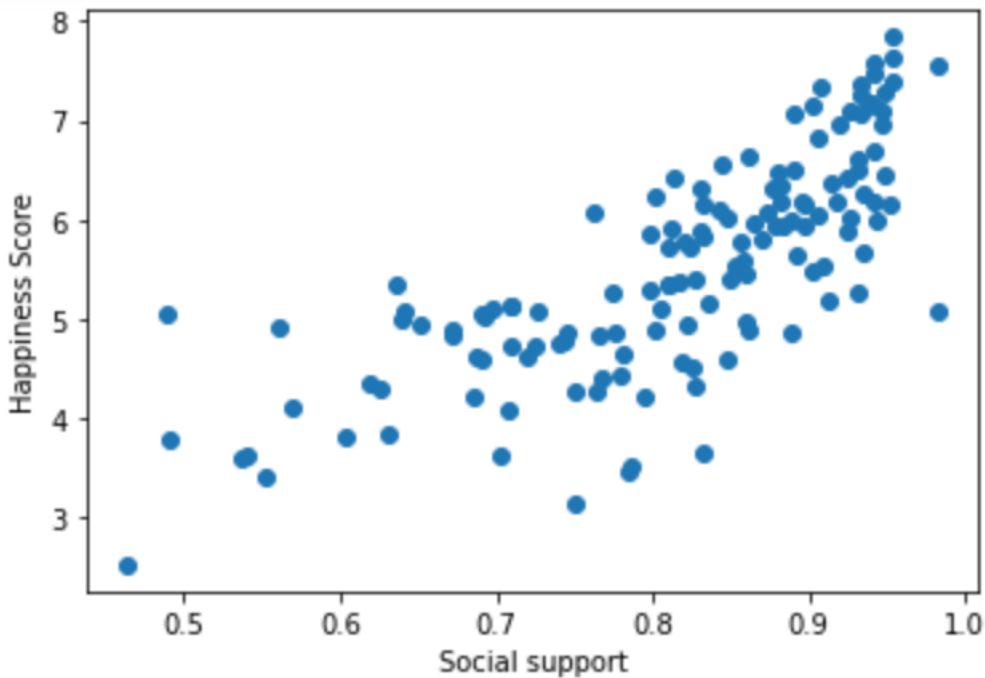
Based on p-values, we see that at a 95% confidence interval, the following variables are statistically significant:

1. norm_Logged GDP per capita
2. norm_Social Support
3. norm_Freedom to make life choices
4. norm_income_per_capita
5. norm_gini

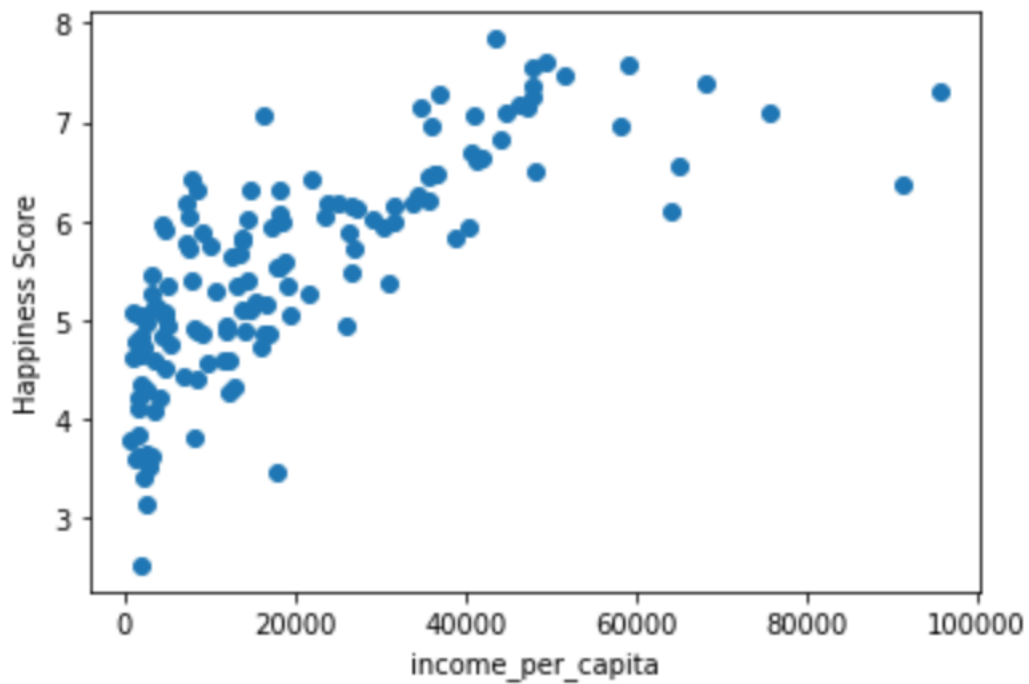
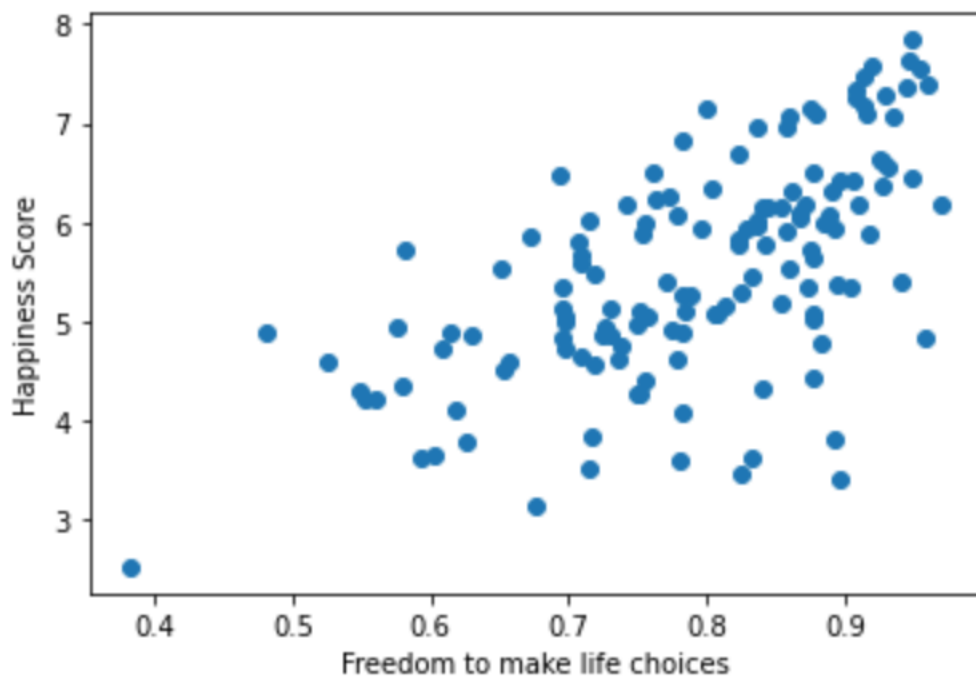
Here we can see the relationship between these variables and the Happiness Score:



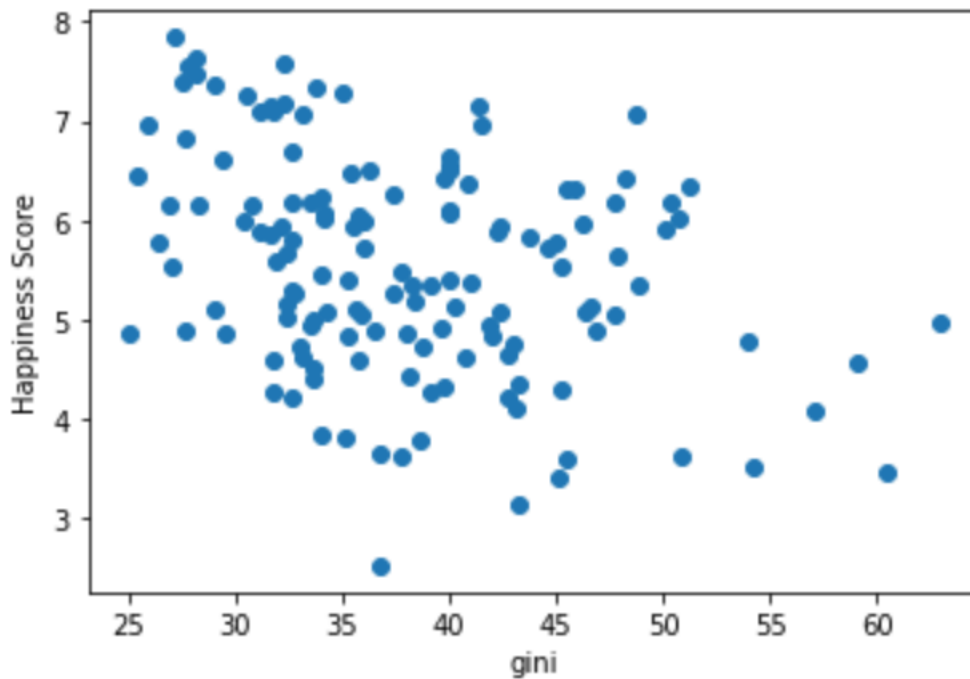
There seems to be a positive correlation between happiness and GDP per capita



Same with social support but the relationship seems a bit nonlinear



It appears that there is a positive correlation between income per capita and Happiness but up to a certain extent. In fact there are a few countries with the highest income per capita but still rank lower than countries with 40k income per capita.



There is a loose negative correlation between the Gini index and happiness score. The higher the index, the more unequal a country so income inequality appears to act as a drag on happiness.

Of the three models that we ran, the one that had the lowest MSE and RMSE was the model with all variables despite having some features that were not statistically significant. From a prediction standpoint we would go with the model with all values however when drawing inference I would only focus on the 5 significant variables.

Afterwards, we used a lasso regression and Grid search to find the best alpha value. For a Lasso regression model using the 5 significant variables plus indicator variables, the best alpha was 0.0075, a very small value making this model very similar to Ordinary least squares regression. However Lasso did yield better predictions on the test set than when using any of the Linear Regression models. The Mean Absolute error, Mean Squared Error and RMSE were all lower than any linear regression model.

For the time being the Lasso regression model is the best one for predicting the happiness score for a country.

	Mean Absolute Error	Mean Squared Error	Root Mean Squared Error
index			
linear regression all vars	0.404218	0.304642	0.551944
linear regression with five features	0.445276	0.348666	0.590480
linear regression with five features and regional indicators	0.403767	0.323307	0.568601
lasso regression with five features and regional indicators	0.392064	0.281414	0.530485

Conclusions

First, we find that the Lasso regression model does a good job of predicting the happiness score with the 5 significant features and regional indicator variables.

This tells us that we can tell a lot about happiness based on the region of the country and its levels of equality, freedom, social support, GDP and Income. So naturally rich and developed countries will rank highest in happiness. Within rich developed countries, those that have high social support and are equal are the ones that do the best. This is why Scandinavian countries rank the highest in happiness while the U.S does not, despite being the richest nation on the planet.

In terms of inference, we the main takeaway is that being rich is important for a country's happiness, if the bounty is distributed unevenly it will result in lower happiness levels. Also, being rich and having a good social safety net translates into better overall happiness for a country and is an advantage of the social democratic model.