

## What Makes A Country's Citizens Happy?

### Introduction

#### *Background*

Happiness in life is something most people want in life, rich or poor. Although there are factors in our individual lives that contribute significantly, for a country as a whole, it is heavily linked with quality of life. However, happiness does not increase linearly with having a higher income, so we will explore other factors that may be related to determining the happiness of a country's citizens.

#### *Problem*

Data that might contribute to predicting whether a country's citizens will be more or less happy may include measures of income, measures of wealth inequality, life expectancy, years of schooling, and alcohol consumption.

#### *Interest*

Individuals would be interested in the insights to how the happiest countries live. City, state, and federal governments might glean some useful information from how to shape their policies to help their citizens stay happy. Relationships between inequality and happiness might motivate lawmakers and help shape economic policies.

### Data

#### *Sources*

Most of the country data and metrics could be found on Kaggle. The median income figures were found on the cgdev.org website. Here are the links:

1. <https://www.cgdev.org/blog/world-bank-poverty-statistics-lack-median-income-data-so-we-filled-gap-ourselves-download-available>
2. <https://www.kaggle.com/unsdsn/world-happiness?select=2015.csv>
3. <https://www.kaggle.com/nitishabharathi/gdp-per-capita-all-countries?select=GDP.csv>
4. <https://www.kaggle.com/kumarajarshi/life-expectancy-who?select=Life+Expectancy+Data.csv>

Kevin Wang  
November 20, 2020

### *Cleaning*

Downloaded data from multiple sources were combined into one table by joining on the country's name. Inner joins were used to exclude missing data. Many tables had multiple years, so I stuck with data around 2012 for most of the datasets, and used 2015 for the happiness score dataset since that was the closest to 2012 I can find. For the median income, the calculation by this organization was based on daily instead of annual. In their example, real median household income for the U.S. was \$53000, translating to \$48 per day per person for a family of 3. I then scaled the daily median income for individuals to annual and to households. While the absolute number does not matter as much, I am more interested in the variance in the ratio of the median annual income to GDP per capita between various countries. The difference in a median measure and a mean measure captures the amount of inequality in wealth (skewing of a relatively normal distribution). Other formatting details were mostly for rounding and clarifying column labels.

### *Feature Selection*

The features that I selected to be of potential relevance in predicting a country's happiness are:

1. Medium Annual Income (Household)
2. GDP Per Capita
3. Ratio of Medium Annual Income / GDP Per Capita
4. Average Life Expectancy
5. Average Years of Schooling
6. Annual Liters of Alcohol Consumed Per Capita

## **Methodology & Exploratory Data Analysis**

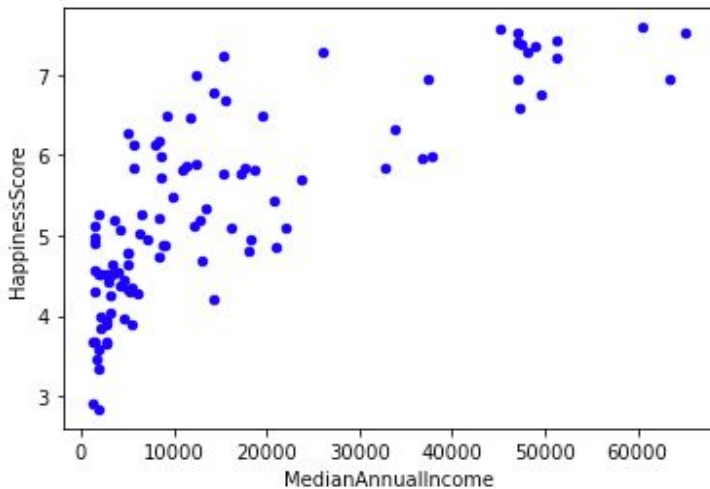
### *The Target Variable - Happiness Score*

The happiness score in this combined dataset was obtained through a Kaggle dataset (#2) from a world happiness survey conducted in 2015. It is simply an average, for each country, of the responses of individuals who indicated that they are a certain degree of happiness from 0 to 10 where 10 is the happiest.

Kevin Wang  
November 20, 2020

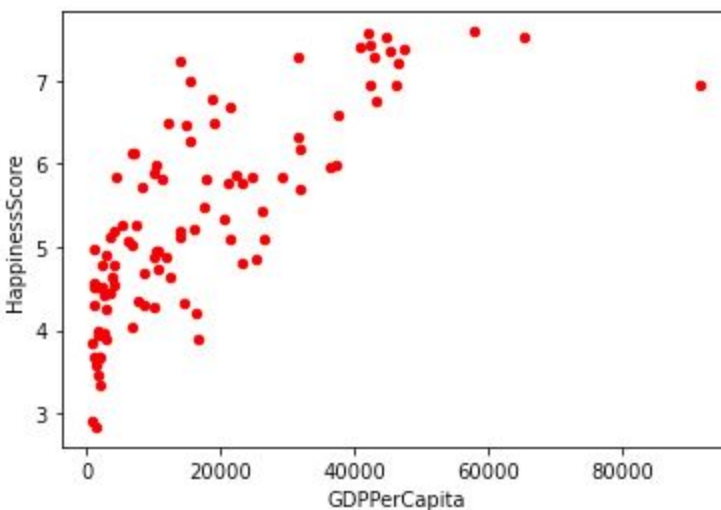
### *Relationship Between Happiness and Median Annual Income (Household)*

As expected, generally higher incomes translates to higher standards of living, which is reflected in happiness scores. We see that happiness improves markedly early on, but after reaching the 30,000 to 40,000 dollars range, additional income does not make a substantial difference.



### *Relationship Between Happiness and GDP Per Capita*

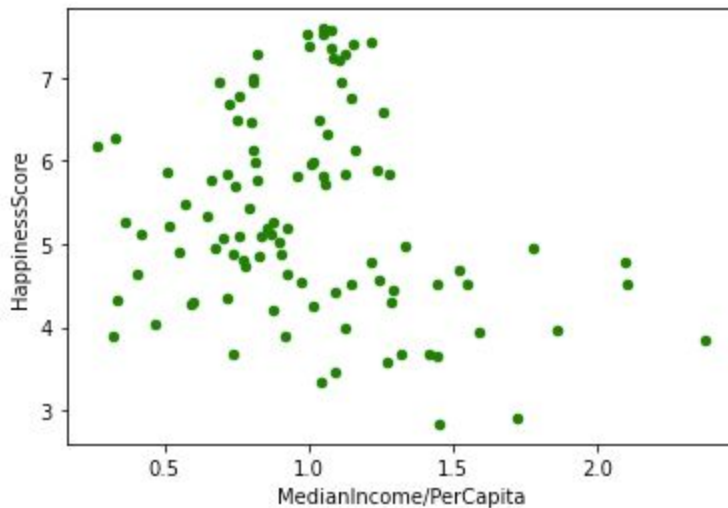
As expected, generally higher GDP per capita, a measure of productivity which is probably roughly reflective of living standards, translates to higher happiness scores. Like median annual income, happiness increases steeply with productivity at first, but levels off beyond the 30,000 to 40,000 dollars range.



Kevin Wang  
November 20, 2020

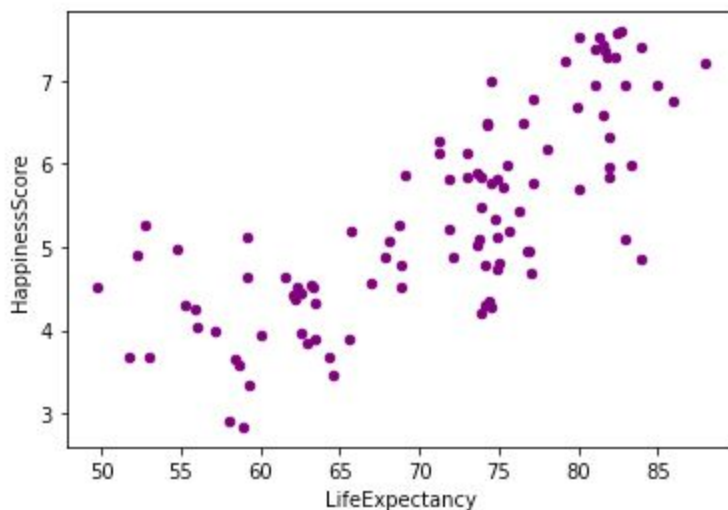
### *Relationship Between Happiness and Ratio of Median Annual Income / GDP Per Capita*

This relationship is a bit more nuanced. Glancing upon the graph, we see that there might not be an obvious linear relationship between the ratio of MAI/GDPPC. Instead, happiness seems to peak around a ratio of 1. This means that we might have to transform the data further to glean more insightful information that can be digested in a linear regression format. In the final analysis, we take the absolute value of the difference between the ratio and 1 to try to see if there's a meaningful relationship there.



### *Relationship Between Happiness and Life Expectancy*

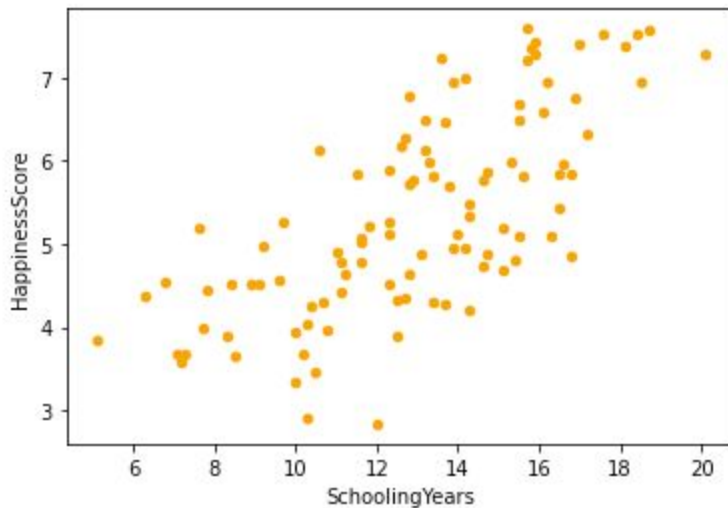
In evaluating the relationship between happiness and life expectancy, higher life expectancies seems to correlate with higher happiness scores. This makes sense since people living longer is generally a good thing, and may also reflect better standards of living during those lifetimes.



Kevin Wang  
November 20, 2020

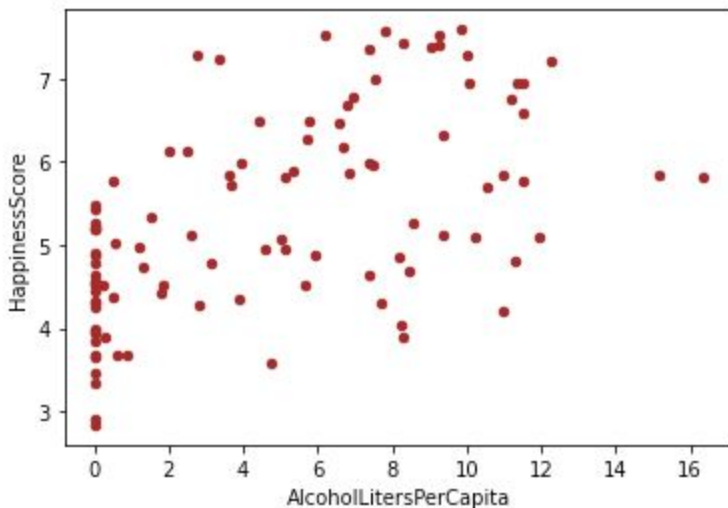
### *Relationship Between Happiness and Schooling Years*

Length of schooling also seems to have a direct positive correlation with happiness. While the connection may seem obvious, this may be due to the fact that better access to education provides more opportunities to individuals throughout their life and can translate to better standards of living and therefore happiness.



### *Relationship Between Happiness and Alcohol Consumption*

The connection between alcohol consumption and happiness is not clear, but at a glance, it seems that moderation is most connected to higher happiness scores while the extremes of alcohol consumption result in lower happiness scores. Because of this pattern, we will transform the data to the absolute value of the difference of alcohol consumption per country and the halfway point (8 liters per capita per year).



## Results & Discussion

After transforming the non obvious data relationships found between happiness and the factors MAI/GDPPC and Alcohol Consumption, a multiple linear regression was run with the following coefficient result set:

- **2.74417262e-05** for Median Annual Income. This matches our initial expectations since there is a positive relationship between happiness and this factor.
- **6.32633004e-06** for GDP Per Capita. This matches our initial expectations since there is a positive relationship between happiness and this factor.
- **-2.22168194e-01** for MAI/GDPPC ratio distance from 1. This makes sense because a ratio of 1 represents the least amount of inequality possible (mean and median are the same, so no skewing towards either direction). The further the distance of the ratio from 1, the lower the happiness score, which means more inequality contributes to lower degrees of happiness.
- **3.95047976e-02** for Life Expectancy. This matches our initial expectations since there is a positive relationship between happiness and this factor.
- **1.07721623e-02** for Schooling Years. This matches our initial expectations since there is a positive relationship between happiness and this factor.
- **-3.63413354e-02** for Alcohol Consumption distance from 8. 8 liters of alcohol consumption per year is the middle number of the data set with the minimum amount consumed being 0 and the maximum being 16. This measure shows that happiness decreases the further away the amount of alcohol consumed gets away from the middle number, with the extremes corresponding to lower happiness scores. This seems to indicate that moderation in alcohol consumption is best. Perhaps it can be explained that too much contributes to alcoholism, poor lifestyle habits, and their impact on life choices. Too little might be reflective of limited personal freedoms or a cultural taboo.

Overall, this exploration of the relationship between happiness and various factors seems to indicate to countries that improving standards of living while keeping inequality low and moderating alcohol consumption are the practices that will most likely keep its citizens happy, which makes sense.

Variance ( $R^2$ ) was calculated to be .69, which is decent, but could be better. More accurate conclusions would probably be more achievable with a larger, more complete dataset of countries, or perhaps even on a local level within countries.

[https://github.com/kevwang19/Coursera\\_Capstone/blob/master/Data%20Science%20Capstone.ipynb](https://github.com/kevwang19/Coursera_Capstone/blob/master/Data%20Science%20Capstone.ipynb)

## Conclusion

This study sheds light on how countries can best govern, and can serve as a general guideline for governments at all levels on how to maintain a thriving community. Further studies that could be of interest can include similar exploration of these relationships at more local levels.