**HYPOTHESIS**

The happiness of citizens in a country depends largely on the prevailing socio-economic conditions in that country and the happiness score, developed by the World Happiness Report, shows how positively people evaluate their lives on the average. However, there are other factors that we believe should prevail for the citizenry of a country to be happier and that the higher the rate these factors present, the better lives people have. Therefore, there must exist a positive relationship between factors such as corruption perception, generosity, GDP per capita, life expectancy, education and human freedom with the extent of happiness in a country.

**RESEARCH QUESTIONS**

* What is the relationship between a country’s happiness score and other conditions such as corruption perception, generosity, GDP per capita, life expectancy, education and human freedom?
* What are the indicators that have the most significance on happiness?
* Are people happier in countries with high GDP per capita than countries with low GDP per capita?
* With the information gathered from regression analysis, what does the highest significant indicator tell us? What about the lowest?
* How does the regions compare to one another?

**Description of Data**

**HAPPINESS SCORE** – The happiness score is an index generated by the World Happiness Report by providing global ranking of countries based on its citizen’s self-reported well-being. The report uses the Gallup world Poll which asks respondents a similar set of questions across more than 160 countries to report on their evaluation of the quality of life they have. The overall happiness score is generated based on these responses and countries are ranked according to their individual scores. The data on happiness used in this research was therefore obtained from the 2019 World Happiness data and not based on our calculations as are the other variables.

Source: <https://worldhappiness.report/ed/2020/cities-and-happiness-a-global-ranking-and-analysis/>

**CORRUPTION PERCEPTIONS INDEX (CPI)** – the index scores and ranks counties around the world based on how corrupt their public sector is perceived to be but unlike the happiness score, the CPI is based on the views of experts and businesses and not the general public. The CPI is calculated using data from thirteen (13) different sources including data from the World Bank, World Economic Forum, Private Risk and Consulting Companies and Think Tanks. The CPI uses a scale of 0 to 100 where zero is highly corrupt and 100 is considered ethical and honest.

Source: [2020 - CPI - Transparency.org](https://www.transparency.org/en/cpi/2020/index/nzl)

**GENEROSITY SCORE** – The index is calculated by the Charity Aid Foundation (CAF) which uses the Gallup World Poll in over 140 countries to gain insight into the giving and charity behavior within these countries. The three (3) main aspects of giving behavior considered include helping strangers, donating money to charity and volunteering time. The CAF World Giving Index Score is calculated as a combined average of the proportion of people who reported one or more of the giving behaviors.

Source: [CAF World Giving Index 10th Edition| Ten years of giving trends (cafonline.org)](https://www.cafonline.org/about-us/publications/2019-publications/caf-world-giving-index-10th-edition)

**Gross Domestic Product (GDP) Per Capita** – GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. GDP per capita is therefore the gross domestic product of a country divided by its midyear population. This value is normally denominated in U.S. dollars as is the case for our dataset.

Source: [World Development Indicators | DataBank (worldbank.org)](https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.PCAP.CD&country=)

**LIFE EXPECTANCY** – This is simply the average period a person living in a particular country may expect to live. This is the average time a person or group of people are expected to live based on the year of birth, current age and other demographic factors such as gender. The data on life expectancy used in this project was obtained from the World Bank which calculates this based on reports from organizations such as the United Nations Population Division, the U.S. Census Bureau and other National Statistical Offices.

Source: [Life expectancy at birth, total (years) | Data (worldbank.org)](https://data.worldbank.org/indicator/SP.DYN.LE00.IN?end=2018&start=1960&view=chart)

**EDUCATION** – Our goal for the education indicator was to obtain data on the completion percentage of secondary schooling for each country. However, we were unable to find a comprehensive dataset that worked with our overall goal for the final data frame. Instead, we choose to use the enrollment rate in secondary schooling to represent the percentage of a countries population who have completed primary school (grades 1-8) and have decided to continue their education. Please note, that some countries will indicate an enrollment rate higher than 100%. Upon further investigation, the enrollment rate focuses on the individuals in a specific age group who should be enrolled in secondary school. Therefore, countries with greater than 100% enrollment shows individuals outside the expected age who have enrolled in secondary school. For example, someone age 25 who did not attend secondary school decided to go back to school to finish their education.

Source: https://data.worldbank.org/indicator/SE.SEC.ENRR

**HUMAN FREEDOM SCORE** – This index score was obtained from the CATO Institute, a public policy research organization. The purpose of the Human Freedom Index is to present a measurable index that can be used to observe relationships between human freedom and other social and economic indicators. The CATO Institute uses 76 distinct indicators of personal and economic freedom to calculate their Human Freedom Index. For the purposes of our research, we have borrowed their final HFI score to add to our data frame in order to understand what role human freedom has in the overall happiness of the world.

Source: https://www.cato.org/human-freedom-index/2020

**DATA EXPLORATION AND CLEAN-UP**

We divided the 6 indicators between ourselves and were each tasked with finding data that would allow us to run our analysis. Most of the raw datasets were available in CSV, XLSX and JSON. Once we found the data we wanted to use, we had to examine and prepare the data for an eventual merge in our main notebook. From each dataset, we wanted to create concise data frames containing Country, Region and a value that best summarized that particular indicator. Those data frames were exported to CSV files and were ready for the final merge in our main Jupyter Notebook.

Our main Jupyter notebook is named consolidated\_data\_df and as stated above, we read in each of the new CSV files created from our initial clean-up of the data. We did experience some issues once we read in all the data to one data frame. We quickly realized that of the 200+ countries that were captured by the data we collected, most did not have data available to fill in values under the 6 indicators. This would reduce our data frame down to 88 countries that had data to run our analysis. The final part of our clean-up process was formatting the GDP column to a smaller number to help visualize the data easier and to format the columns to 2 decimal places.

**DATA ANALYSIS**

For the analysis, we first wanted to find out what, if any, relationship existed between the indicators and the Happiness Score. We achieved this by plotting each of the six indicators individually against the Happiness Score with a scatter plot. We then used the Scipy.Stats library to run the linear regression on each graph. As we expected, a positive relationship existed between each indicator and the Happiness Score. Our goal now was to examine each slope coefficient and r-value to determine which indicator has the most impact on happiness score. What we discovered was that GDP per capita and Life Expectancy have the highest R-values, which could suggest that amongst the 6 factors, they are the most important factors affecting the happiness level in a particular country.

Now that we have determined which factors are more significant than the rest, we decided to examine the countries that registered high in Human Freedom Score and Life Expectancy. One way we analyzed the relationship between Life Expectancy, Human Freedom, and Happiness Score was through a categorized bubble chart. We set the x-axis as Life Expectancy, the y-axis a Human Freedom Score, and the bubble size to happiness score. The chart clearly depicts the relationships between the two most significant indicators and happiness score. The higher the Life Expectancy and Human Freedom Scores, the larger the bubble (Happiness Score). A couple examples of this relationship are Switzerland and Zimbabwe. Switzerland has a Life Expectancy of 83.4 and a Human Freedom Score of 8.82. Accordingly, their Happiness Score is one of the highest at 7.48. Zimbabwe, on the other hand, has some on of the lowest Life Expectancies at 60.7 and a low Human Freedom score of 5.59. Their Happiness Score is a 3.663. To contrast this comparison, we also plotted a second visualization that shows the relationship between CPI Score, Generosity Score, and Happiness Score, which have two of the lowest r-value and slope coefficients. This visual demonstrates how low corruption and a generous population has a less significant impact on happiness. One example of this is Botswana. They have low corruption and a Generosity Score that is closer to the median, however, their happiness score is low relative to other countries at 3.488. Mexico, on the other hand, has a high Happiness Score at 6.595, but their Generosity and CPI scores are low relative to other countries at 28 and 29 respectively.

**CONCLUSION - Revisiting the questions we asked above, what have we learned?**

Our conclusions from our data analysis are as follows. Life expectancy and GDP per Capita are the two indicators that most significantly impact happiness. They both have strong positive relationships with Happiness Score and the R-values closest to 1. In regard to region, Western Europe & the European Union have some of the highest Human Freedom Scores, Life Expectancy, GDP per Capita, and Happiness Scores while Sub-Saharan Africa and the Asia Pacific region have some of the lowest values for those indicators and therefore lower Happiness Scores. This is illustrated in our bubble chart entitled “Life Expectancy, Human Freedom, and Happiness.” The X-Axis is Life Expectancy, the Y-Axis is Human Freedom Score, and the bubble size is Happiness Score. As you go from left to right, and bottom to top, the bubble size and Happiness Score increase. The Western Europe region is depicted in green at the top right and the Sub-Saharan Africa region is depicted in yellow towards the bottom left of the chart. Singapore is an outlier in the Asia Pacific region. They have a higher Human Freedom Score and Life Expectancy than the rest of the region and because of this, have a Happiness Score almost identical to a Western European Nation in Spain.

While all indicators were positively correlated to Happiness Score to some extent, we found that Generosity Score had the weakest relationship with Happiness. On our bubble chart entitled “CPI, Generosity, and Happiness,” you can see that Myanmar has one of the highest Generosity Scores at 58, but a relatively low Happiness Score in 4.36. In contrast, Mexico has a low Generosity Score of 29, but a relatively high Happiness Score of 6.59. We also found that Generosity Score was less region-specific than the other indicators. While each region’s values were clustered together for most other indicators, the values for the countries within each region for Generosity had a larger variance between minimum and maximum scores in each region. And lastly, the top five countries according to Happiness Score for 2019 were Finland, Denmark, Switzerland, Sweden, and Austria.