Exploring Average Global Happiness and the Affecting Variables

Group 11: Amanda Maguire, Omar Altaher, Anace Alhashhoush, and Ali Abughazaleh

Department of Information Management and Business Analytics, Montclair State University

INFO584: DATA VISUALIZATION

Dr. Eyyub Kibis

Feb 25, 2023

Exploring Average Global Happiness and the Affecting Variables

Abstract:

Through analysis, visualizations, and a written report we aim to explore the relationship between various variables and a country's happiness score. Analyzing and identifying potential correlations between the various factors. The data was gathered and refined from the Kaggle dataset "World Happiness," which provides happiness scores and rankings based on data from the Gallup World Poll. While the variables in question are mostly emotion based, a prominent theme was the effect of GDP, a monetary value, on happiness score. As a team, we aimed to discover the correlation between the variables driving happiness. An essential question seen time and time again is “Can money buy happiness?” From our data exploration, we can confidently say it may not buy it, but it definitely helps. The social and emotional factors contributed to the happiness level but none of them were as highly correlated as GDP. Happiness as the dependent variable is defined as a state of well-being and contentment. Through the analysis, visualizations, and written report you will find that the state of happiness scored much higher as GDP emerged amongst the countries.

Introduction:

Our project will focus on several key questions to establish relationships between the variables contributing to happiness. To arrive at our answers, we first cleaned, formatted, and conjoined 5 years of data, to have all the same variable types and remove any null entries. The data includes up to 157 countries per each year from 2015-’19, including all special and unique demographics. With the different backgrounds in mind, we must test the linearity of the data. Do the variables have the same effect across each country, or similar, proving linearity? Narrowing in on how a country's happiness score compares to its GDP per capita? Determining on how the increase of GDP controls the happiness by majority, but how do the remaining variables factor with the GDP? GDP being the only monetary variable amongst more emotional factors resulting in a higher happiness score.

As mentioned, the countries come from all over the world with unique characteristics and demographics. A goal would be to have all countries be a contender for the #1 happiness score. Unfortunately, that is not the case and patterns will be analyzed to understand why the countries in the top remain and vice versa for the bottom of the list. The main goal being global happiness, we set out to reach that objective, by developing SMART goals.

Our first smart goal we set out to achieve is to increase happiness scores by 20% within 3 years, in order to improve the happiness score in a certain country. At first glance, every year has the same top 10 happy countries. With the country’s having all different backgrounds the variables contributing remain the same. Focusing on the impacting variables highly correlated with the happiness score, we explore the patterns to see how we can accelerate the happiness score of a low scoring country. The patterns and trends spotted from happier countries will give great guidance to improving others. Within the years’ time, if the correlated variables are improved, the overall happiness score should rise, resulting closer to our goal of global

happiness. We don't need to dim one country's candle to have one shine brighter, but we can see how their flame was sparked, sparking another.

A key variable contributing is GDP, leading to the next SMART goal. Aiming to increase the country's GDP per capita by 10% within 2 years, in order to improve its happiness score. Globally, many countries are still developing economically compared to others. As GDP rises, happiness score and rank increase. If the country's government focused on its economic well-being, without abandoning the other variables, it will find its community in an improved state of happiness. The main sources of economic growth are increasing the size of the workforce and growth in productivity. If the country has a certain good that can be exported, finding the right importer can increase both the jobs needed and the opportunity. For instance, the top country, Norway, has one of the largest reserves of seafood, hydro-power, lumber, minerals, natural gas, and freshwater. This supports the two main sources of economic growth, as they have an extremely low unemployment rate. The increase in GDP implements the country being fueled economically which results in global improvement.

Choosing intellectual and attainable goals, we aim to use various data visualization techniques to present the data in an interactive and easy to understand format. The visual representation should provide valuable insights and information pertaining to global happiness. Using the dashboard created, it allows the user to explore and help restore the data too work towards global happiness.

Data Set:

The World Happiness Report dataset was collected by the Sustainable Development Solutions Network (SDSN) in collaboration with Gallup, with the purpose of providing information on happiness scores and various factors that contribute to happiness. The data were collected annually from a representative sample of individuals in each country, varying sample sizes by country. The Cantril ladder question was used to assess happiness levels, where individuals were asked to rate their current life on a scale from 0 to 10. Additional survey questions were used to collect data on variables such as GDP per capita, social support, and freedom to make life choices.

The data we worked on is specifically for the years between 2015 and 2019, and we conducted data cleaning to ensure coherence across all years. As part of this process, we removed some columns from the individual CSV files to ensure consistency across all years of data. After cleaning the data to ensure coherence, we conducted various statistical tests on the dataset to gain insights into the relationships between different variables and happiness scores.

We visualized the data using various tools such as scatter plots, maps, and bar graphs to identify patterns and trends. These techniques helped us to better understand the factors that contribute to happiness and to explore the relationships between various variables. Data exploration, transformations, and imputations were not heavily conducted since we purposely choose this dataset based on the cleanliness of the data which helped ensure the reliability and

accuracy of our study and allowed us to gain meaningful insights into the factors that contribute to happiness.

Variable	**Data Type**	**Description**
Country name	Categorical	Name of country
Happiness score	Numeric	A metric measured by asking the sampled people the question: "How would you rate your happiness on a scale of 0 to 10 where 10 is the happiest."(Cantril ladder question)
Economy (GDP per Capita)	Numeric	Natural logarithm of GDP per capita
Social support	Numeric	Perceived social support based on the Cantril ladder question
Healthy life expectancy	Numeric	Life expectancy based on responses to the Cantril ladder question
Freedom to make life choices	Numeric	Perceived freedom to make life choices based on the Cantril ladder question
Perceptions of corruption	Numeric	Perceived level of corruption based on responses to the Cantril ladder question
Generosity	Numeric	Perceived generosity based on responses to the Cantril ladder question

Analysis & Results:

To approach our objective, we first conducted a correlation analysis in Excel to see the relationship between our predictor variables and our target variable “Happiness score”.

	Happiness Score	*Happiness Rank*	*Year*	*Economy (GDP per Capita)*	*Family*	*Health (Life Expectancy)*	*Freedom*	*Trust (Government Corruption)*	*Generosity*
Happiness Score	1								
Happiness Rank	-0.992066495	1							
Year	0.007064687	-0.007767823	1						
Economy (GDP per Capita)	0.789283997	-0.794790645	0.019768356	1					
Family	0.648799345	-0.644841509	0.36743122	0.585965534	1				
Health (Life Expectancy)	0.742455742	-0.743654598	0.130302255	0.784337571	0.572650262	1			
Freedom	0.551258003	-0.537942231	0.010353162	0.340510988	0.420360835	0.34074513	1		
Trust (Government Corruption)	0.398020158	-0.372370193	-0.122228663	0.304640109	0.126337959	0.25050215	0.459394015	1	
Generosity	0.137577734	-0.117713311	-0.192587432	-0.014560482	-0.037261611	0.01063811	0.290705511	0.318906512	1

Exploring Average Global Happiness and the Affecting Variables

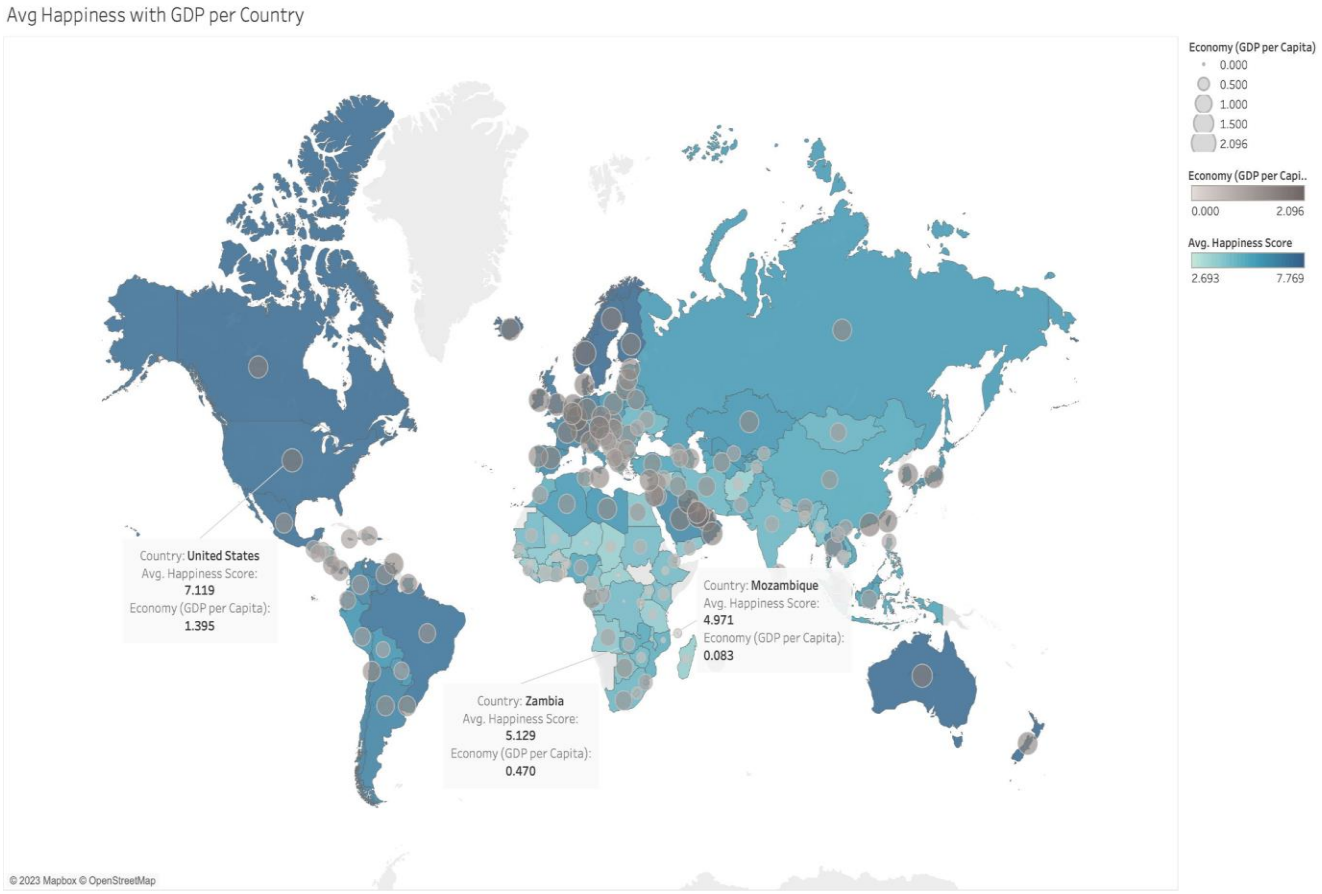
The correlation table above indicates that GDP has the strongest correlation with happiness score. However, it's important to note that correlation does not imply causation. We can note that family, health, and freedom also have a positive correlation with happiness score.

Learning that the target variable and the other variables are highly to moderately correlated, shaped our future analyses. We tend to test the relationships, degree of linearity, and direction by visualizing the correlation using a scatter plot with a trend line visual. Using Tableau, allows us to evaluate how well the model fits the data. Note: We placed the average happiness score per year on the X- axis and the other variables on the Y-axis.

As displayed in the scatter plot, we can see that the visualization translates the relationships, degree of linearity, and direction of data in the correlation table above. As a result, all variables have a positive relationship with the average happiness score. On the other hand, the linearity degree varies among the variables, however, we can observe a high degree of linearity in the case of average GDP, family, health, and freedom.

To discover additional findings, we used Tableau to construct the map below with a dual axis to identify if average happiness and GDP are positively correlated in a bigger picture. If this is not the case, then other factors may influence happiness. Note: Hue is used to showing average happiness score, with darker colors indicating higher happiness scores, and circle size to show GDP.

Exploring Average Global Happiness and the Affecting Variables

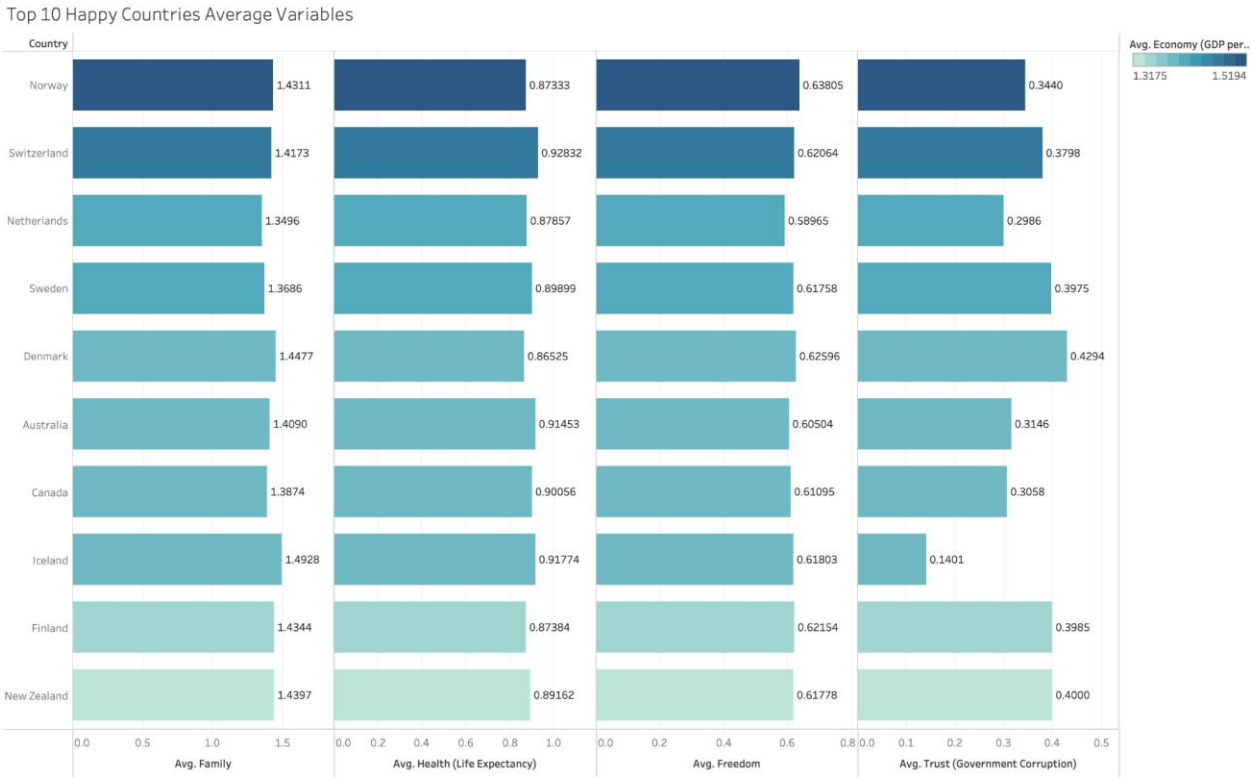


We noted from the map above that in most cases, countries with low happiness scores are in Asia, Africa, and some countries in South America. On the other hand, high average happiness score and GDP take place in Europe, North America, and Australia in most cases. The map displays countries located in Asia, Africa, and some countries in South America with high GDP, but yet less happy. Contrasting, GDP may also be affected by factors other than happiness.

By visualizing the differences in happiness score and GDP across countries, we can identify patterns that helped us better understand the factors influencing happiness score. This proved that average GDP and average happiness score may not always go together. For instance, the case in countries such as Zambia and Mozambique show that they almost have equal value in terms of average happiness, but not in average GDP. This opens up a chance to have other factors, such as family, freedom, health, and government trust or corruption level play a role in a country's happiness.

After recognizing that other factors may affect the average GDP, we constructed a comparative bar chart in Tableau. The chart compares the average values of family, health, freedom, and trust across the top 10 happy countries, using average GDP as the hue. Our analysis revealed that for a country to have a high happiness score, it must have a combination of multiple

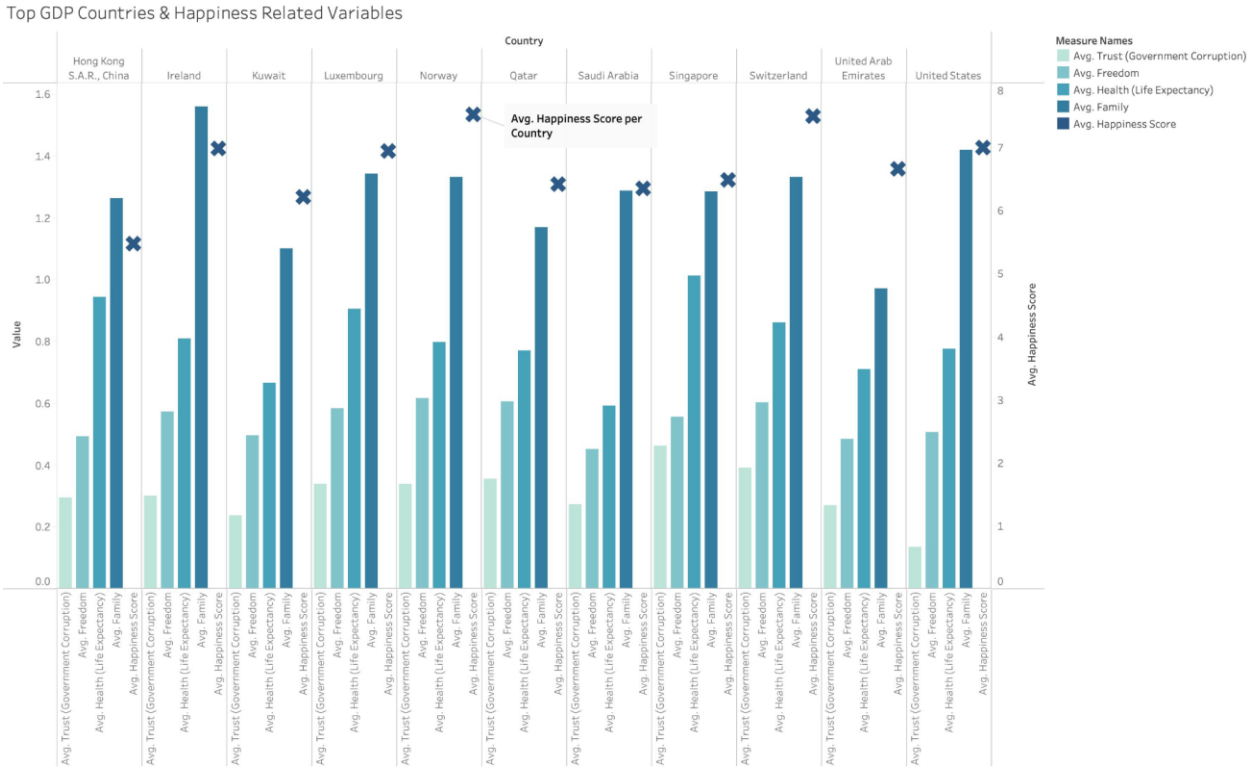
factors, not just high average GDP. Average family, health, freedom, and trust were all strongly associated with happiness score. Note: Hue is being used to portray the average GDP, with darker colors indicating higher GDP, for the Top 10 Happy Countries.



This visualization helped us identify commonalities among the top 10 happy countries and get a more nuanced understanding of the factors that contribute to happiness. We noted that a high happiness score average is related to the other independent variables. In other words, for a country to be happy it must have a combination of all the factors, not just high GDP average, even though it's the highest correlated.

To discover more details about how the variables may vary in affecting happiness score other than GDP, we constructed a bar chart. This will enable us to point out the game changing variables in terms of happiness along with high GDP per capita. The bar chart below displays the highest GDP countries comparing happiness score, family, health, freedom, and government trust. Note: Hue is used to show the independent variable types, for the Top GDP Countries per capita in correlation to average happiness score.

Exploring Average Global Happiness and the Affecting Variables



The chart above indicates several observations. Firstly, all high GDP countries have average family followed by average health, freedom, and trust in that order. This ranks the priorities and weight of each variable which results in a country having a high GDP which follows our initial correlation study. Secondly, as mentioned previously, happiness does vary among each country, even for the highest top GDP countries. Finally, some countries' average happiness score may be less than other ones' depending on impacting factors.

Once we analyzed our visualizations, we reverted back to our correlation matrix. The visualizations played a key role in identifying trends and patterns pertaining to global happiness scores. To ensure that we can use our findings to answer our research questions and accomplish our SMART goals, we used correlation methods. The predictive analysis enabled us to conclude that we can reach our goals of obtaining high global happiness scores across all countries according to the resources of each country. Below the visuals contain the methodology used to create predictive models based on the variables included within the data set.

Exploring Average Global Happiness and the Affecting Variables

Stepwise Regression Analysis
Table of Results for Backward Elimination

Model with all variables entered.

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	758.5177	#########	#########	0.0000
Residual	775	234.2567	0.3023		
Total	781	992.7745			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2.1775	0.0798	27.2826	0.0000	2.0208	2.3342
Economy (GDP per Capita)	1.1468	0.0828	13.8568	0.0000	0.9843	1.3092
Family	0.6411	0.0806	7.9523	0.0000	0.4828	0.7993
Health (Life Expectancy)	1.0039	0.1314	7.6405	0.0000	0.7460	1.2619
Freedom	1.4791	0.1634	9.0534	0.0000	1.1584	1.7999
Trust (Government Corruption)	0.8536	0.2233	3.8232	0.0001	0.4153	1.2919
N Generosity	0.5936	0.1756	3.3797	0.0008	0.2488	0.9384

Happiness Score = 2.1775 + (1.1468 * Economy) + (0.6411 * Family) + (1.0039 * Health) + (1.4791 * Freedom) + (0.8536 * Trust) + (0.5936 * Generosity)

Freedom removed.

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	85.4420	21.3605	375.9574	0.0000
Residual	777	44.1462	0.0568		
Total	781	129.5882			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.0166	0.0339	-0.4886	0.6253	-0.0831	0.0500
Family	0.2514	0.0316	7.9509	0.0000	0.1893	0.3134
Health (Life Expectancy)	1.0397	0.0430	24.1953	0.0000	0.9553	1.1240
Trust (Government Corruption)	0.5341	0.0881	6.0596	0.0000	0.3611	0.7072
Generosity	-0.1930	0.0739	-2.6124	0.0092	-0.3380	-0.0480

No other variables could be removed from the model. Stepwise ends.

GDP per Capita = -0.0166 + 0.2514(Family) + 1.0397(Health) + 0.5341(Trust) - 0.1930(Generosity)

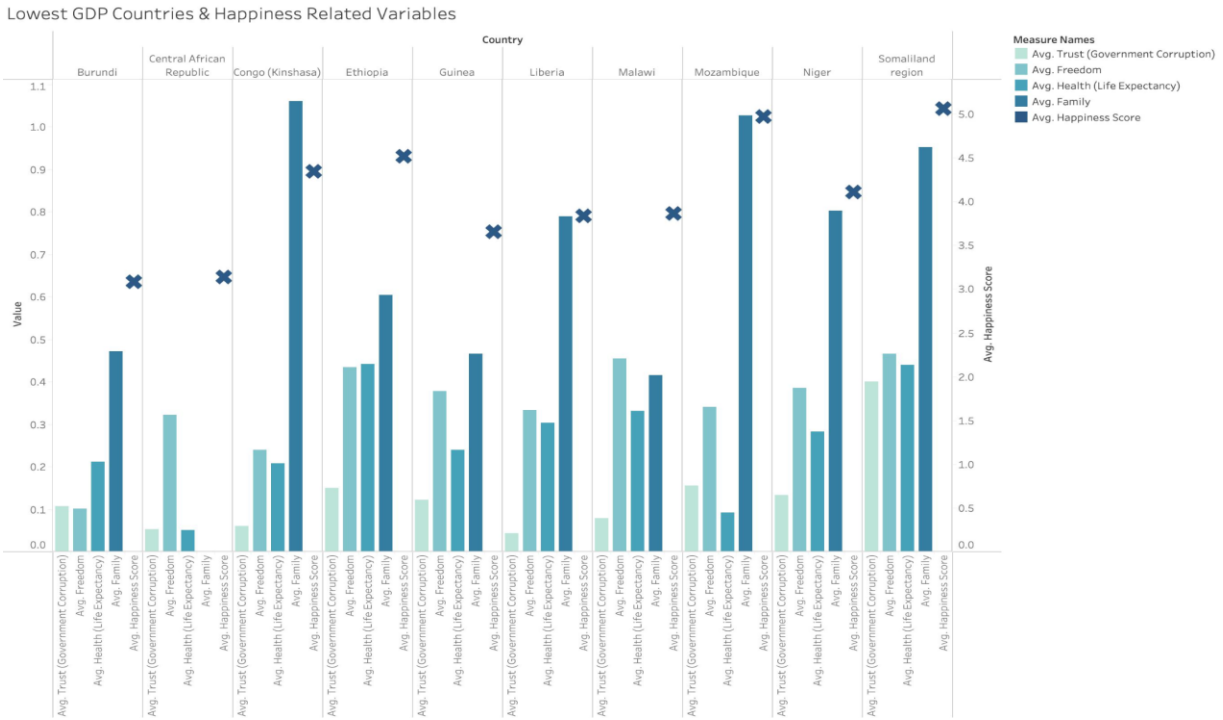
With these formulas, we can estimate the happiness score or GDP given the variables of the data set. This is extremely impactful for concluding our SMART goals as well as estimating global happiness.

Limitations:

The data collected was impactful for gaining insights and making informed decisions on "Exploring Average Global Happiness and the Affecting Variables", but it is important to note the limitations within the data set. Throughout the analysis, we focused on the average variables

affecting happiness within each year as well as focused on the top 10 countries within different categories (GDP and happiness score). This process allowed us to make solid conclusions and connections between variables. We used average scores to overcome the missing continuity of countries in some years. We proceeded with different categories to explain our claims, answering the questions and goals we set out to prove. Within the 5 years time, the data were collected, some countries were still developing and innovating, while others were much further along. If we had the data including 2020-2022, we can assume that the countries' GDP would have risen, resulting in a potentially higher happiness score. On the other hand, during those years, health due to Corona Virus would have been much lower, resulting in lower happiness scores and GDP. Without the data, we cannot conclude which way the pendulum would swing for happiness within those years. We can only comment and make predictions based on the variables included in the data set while recognizing that there may be factors highly correlated but not captured or considered within the data. For example, in the scatterplot visual, we depicted a low correlation degree with the average trust. Concluding that the scatter plot model fits variables such as GDP, family, and health, more than other variables such as freedom and trust with some outliers.

While for the below, we examined the Lowest GDP countries and the effect of the variables. This chart proves that there can be an effect of the relationship of variables based on outside factors. Note: Hue is being used to show the independent variable types, for the Lowest GDP Countries per capita in correlation to the average happiness score.



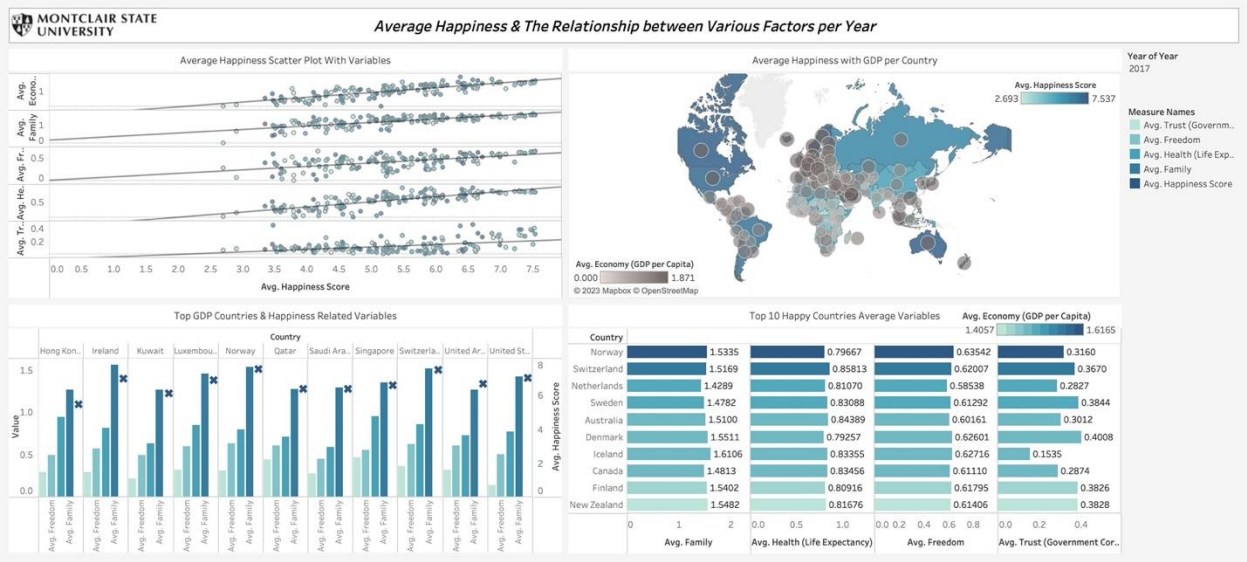
Exploring Average Global Happiness and the Affecting Variables

Conclusion:

To conclude our analysis, there are numerous factors that can influence a country's happiness score. We explored and examined the data with a focus on GDP being a prominent independent variable impacting our dependent variable, happiness score. Through the analysis of our data set, we noted limitations but were still able to find linearity between the variables. We were able to draw conclusions and prove that we can reach our SMART goals, provided with the data. For a country to reach its peak happiness score it must first consider its GDP. As a whole, the remaining factors and variables all play important roles to determine happiness. This is depicted in the visuals provided and explained, while easily usable in our interactive dashboard. From the dashboard, you will be able to make informed conclusions about the country's happiness score given the variables.

Link to our full dashboard:

https://public.tableau.com/app/profile/omar.altaher/viz/AverageHappiness-Global-1/Dashboard2?publish=yes



Exploring Average Global Happiness and the Affecting Variables

Works Cited

Nordland. “Top 5 Reasons Why Norway Has a High Standard of Living.” Free Tour Oslo, 5 Feb. 2023, https://freetouroslo.com/top-5-reasons-why-norway-has-a-high-standard-of-living/#:~:text=It%20also%20has%20one%20of,it%20an%20extremely%20productive%20country.

| By Chad Stone, et al. “Economic Growth: Causes, Benefits, and Current Limits.” Center on Budget and Policy Priorities, 27 Apr. 2017, https://www.cbpp.org/research/economy/economic-growth-causes-benefits-and-current-limits.