A report on how to increase happiness in countries for the Director of Happiness at the United Nations

By Maria Velazquez Marin,

Data Analyst at the United Nations

After a thorough analysis, our recommendation is to focus on improving four key indicators for maximizing national happiness: Logged GDP per capita, Social Support, Healthy Life Expectancy, and Perceptions of corruption. These variables exhibit significant influence on overall well-being. The analysis presented in the report indicates that in order to maximize the countries' overall happiness, countries should have a logged GDP per capita of 11.65, a social support index of 0.98, a healthy life expectancy of 74.7 years, and a perception of corruption score of 0.18.

By prioritizing initiatives that enhance economic prosperity, social cohesion, health outcomes, and combat corruption, we aim to achieve a happiness rating ranging between 7 and 9 with 95% confidence, indicating a high degree of certainty in the expected outcome. This strategic approach addresses the core determinants identified in our model, ensuring a targeted and effective pathway towards greater well-being for the countries.

In our pursuit of national happiness, we struggle with a range of challenges. Factors like economic disparities, social support, health, and perceptions of corruption all play crucial roles in shaping the well-being of our citizens. To optimize the overall well-being and happiness of our nations, it is important that we set our sights on specific key performance indicators. These encompass a logged GDP per capita of 11.65, signifying a sustainable economic foundation for all citizens. A social support index of 0.98 underscores the significance of robust social networks and communal bonds, playing a crucial role in nurturing a sense of belonging and unity within our societies. Moreover, a healthy life expectancy of 74.7 years underscores the need for accessible healthcare and wellness programs, ensuring that individuals can lead fulfilling and active lives. Additionally, a perception of corruption score of 0.179 highlights the significance of transparent governance and accountability in upholding the trust of our citizens. If these conditions are met, we aim to achieve a national happiness rating ranging between 7 and 9 with 95% confidence.

This recommendation comes from a carefully constructed regression model that we meticulously developed. This model serves as a powerful tool in illuminating the important relationship between happiness and several crucial variables. These, including logged GDP per capita, social support index, healthy life expectancy, and perception of corruption, play fundamental roles in determining the overall well-being and happiness of a nation's population. The coefficients we have in our model provide valuable insights into what really drives happiness. For instance, for a one-unit increase in the logged GDP per capita, we expect an increase in the happiness score by approximately 0.238 units, which means that a higher GDP per capita means more happiness. Similarly, when people feel supported by their community, their happiness levels tend to rise: for a one-unit increase in social support, we expect an increase in the happiness score by approximately 1.979 units. A longer healthy life also contributes to overall happiness because for a one-unit increase in healthy life expectancy, we expect an increase in the happiness score of approximately 0.061 units. Conversely, high levels of corruption have a negative impact on happiness: for a one-unit increase in perceptions of corruption, we expect a decrease in the dependent variable by approximately 1.594 units. It's like a barrier that prevents happiness from increasing. By focusing on these key factors, we can work towards creating happier and more fulfilling lives for everyone in our countries. For a more detailed analysis and interpretation of the model, please refer to Appendix A.

The model demonstrates outstanding performance. Based on a dataset comprising 78 observations, our analysis is firmly grounded, offering a strong basis for our

recommendations. It explains approximately 81% of the variability in our response variable, indicating its reliability in capturing the key factors influencing national well-being. We have high confidence that each of our predictors plays a significant role in determining the overall happiness levels. This confidence comes from carefully looking at the data and the numbers associated with each factor. Thus, we feel very confident in using this model to understand and make recommendations about national well-being. For a more detailed analysis, please refer to Appendix B, where additional insights into our comprehensive assessment can be found.

The process I used involved a systematic approach of variable elimination and careful consideration of correlations to arrive at an optimal model. The model's development began with a specific dataset comprising 10 variables in addition to the response variable. The unique approach taken involved an initial regression analysis, followed by a process of selectively removing variables that were identified as less influential factors. Additionally, in cases where variables showed high correlation, I experimented with removing either one to assess its impact on model performance. In this case, it didn't make sense to combine variables because they represented fundamentally distinct aspects of a country's well-being. While all the factors contribute to a nation's overall well-being, they operate in separate spheres. Attempting to combine them could potentially lead to a loss of important information. Thus, by treating them as independent variables, we gain a more accurate and detailed understanding of the diverse factors influencing happiness levels in a country. This process ultimately led to the identification of four variables with low p-values, indicating their significant impact on the model. These variables are Logged GDP per capita, Social Support, Healthy Life Expectancy, and Perceptions of corruption. For a more detailed breakdown of this process, please refer to Appendix C.

Our initial dataset consisted of ten crucial variables, ranging from economic indicators like Logged GDP per capita to social factors such as Healthy life expectancy. This comprehensive data enables a holistic understanding of the determinants of national happiness. For a detailed exploration of the data and its analysis, please refer to Appendix D.

We gathered our data from various reputable websites: World Happiness Report for variables like Logged GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity, and Perceptions of corruption (https://www.kaggle.com/datasets/unsdsn/world-happiness); Unemployment Rate over the past 31 years (https://www.kaggle.com/datasets/pantanjali/unemployment-dataset); Education data focusing on Barro-Lee: Percentage of population age 15+ with tertiary schooling and Completed Tertiary (https://ourworldindata.org/global-

education#literacy); Environmental Performance Index (EPI) Score by country (https://epi.yale.edu/epi-results/2022/component/epi); and Wealth Inequality using the Gini Index (https://worldpopulationreview.com/country-rankings/wealth-inequality-by-country). For more detailed information about the data, please refer to Appendix E.

Appendices

Appendix A: Model and Interpretation

Happiness Score = -1.045668073 + (0.237840271 * Logged GDP per capita) + (1.978795252 * Social support) + (0.061019268 * Healthy life expectancy) - (1.594350402 * Perceptions of corruption)

Coefficient Analysis:

- The coefficient for Logged GDP per capita is 0.2378. This means that for a one-unit increase in the logged GDP per capita, we expect an increase in the happiness score by approximately 0.238 units, holding other variables constant.
- The coefficient for Social Support is 1.9788. This implies that for a one-unit increase in social support, we expect an increase in the happiness score by approximately 1.979 units, holding other variables constant.
- The coefficient for Healthy Life Expectancy is 0.0610. This means that for a one-unit increase in healthy life expectancy, we expect an increase in the happiness score by approximately 0.061 units, holding other variables constant.
- The coefficient for Perceptions of corruption is -1.5944. This indicates that for a one-unit increase in perceptions of corruption (meaning an increase in perceived corruption), we expect a decrease in the happiness score by approximately 1.594 units, holding other variables constant.

Impact Analysis:

	Coefficient	Min	Max	Range	Impact
Logged GDP	0.23784	6.958	11.647	4.689	1.11523303
Social Support	1.9788	0.489	0.983	0.494	0.97752485
Healthy Life Expectancy	0.06102	48.7	74.7	26	1.58650097
Perception of corruption	1.59435	0.179	0.939	0.76	1.21170631

From this table, we see that the variable that has the highest impact in happiness is Healthy Life Expectancy, which can make a difference of 1.58 in the happiness score. Perception of corruption and Logged GDP per capita can make a difference of 1.21 and 1.11, respectively. Finally, the variable with the lowest impact is social support, which can make a difference of 0.97 in the happiness score.

	Coefficients	Optimal
Intercept	-1.04566807	1
Logged GDP per capita	0.23784027	11.647
Social support	1.97879525	0.983
Healthy life expectancy	0.06101927	74.7
Perceptions of corruption	-1.5943504	0.179
	Point Prediction	7.942364
	Interval Prediction	7.000967
		8.883761

In this case, the point prediction is approximately 7.94. However, the probability of being wrong is 1, because there's a very high possibility that the actual value could differ from our point prediction.

The interval prediction is the range of values within which we expect the true happiness score to fall. In this case, we can state with 95% confidence that the true happiness score is likely to be between 7.00 and 8.88.

Appendix B: Model Statistical Analysis

Regression Statistics	
Multiple R	0.905213
R Square	0.81941
Adjusted R Square	0.809515
Standard Error	0.470699
Observations	78

In this case, a Multiple R of approximately 0.905 suggests a strong positive linear relationship between the independent variables and the happiness score.

An R-Square of approximately 0.819 indicates that around 81.9% of the variance in the happiness score can be explained by the independent variables in your model.

The Adjusted R-Square of approximately 0.809 takes into account the number of predictors and provides a slightly more conservative estimate of the model's explanatory power.

A lower standard error (in this case, approximately 0.471) indicates that the model's predictions are relatively close to the actual values.

Finally, we can see that in this case, there are 78 observations.

ANOVA

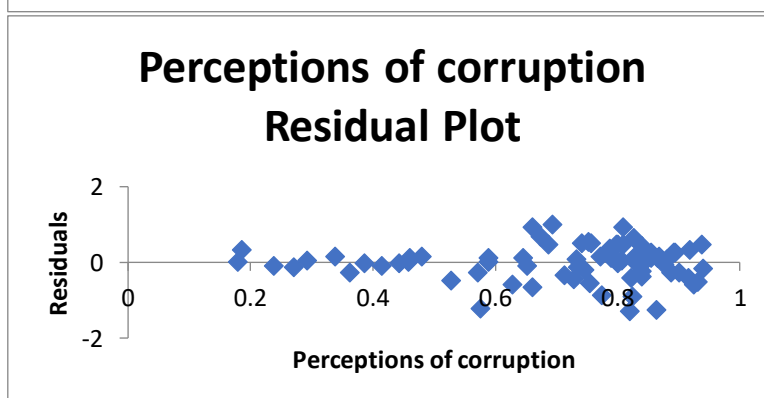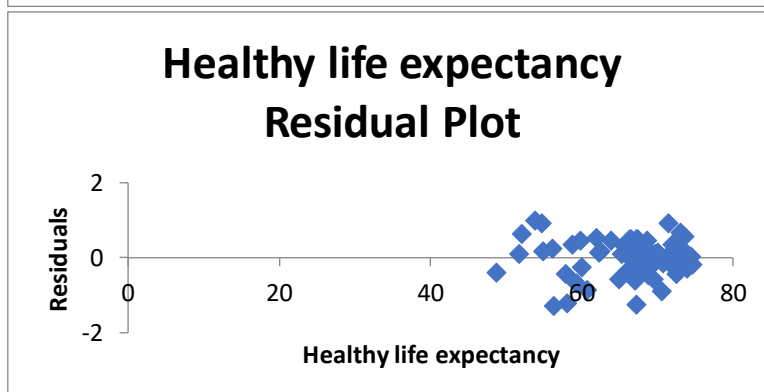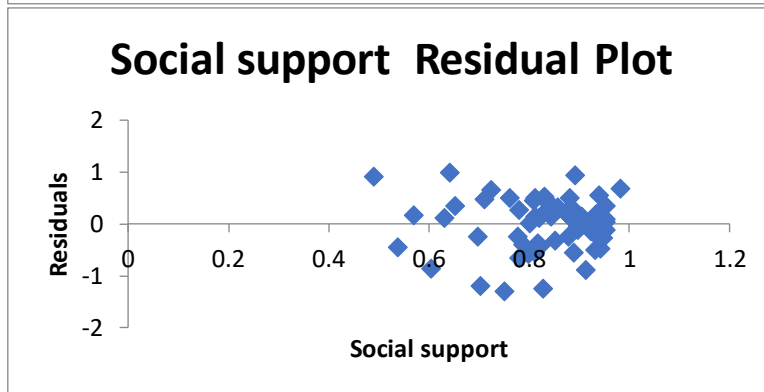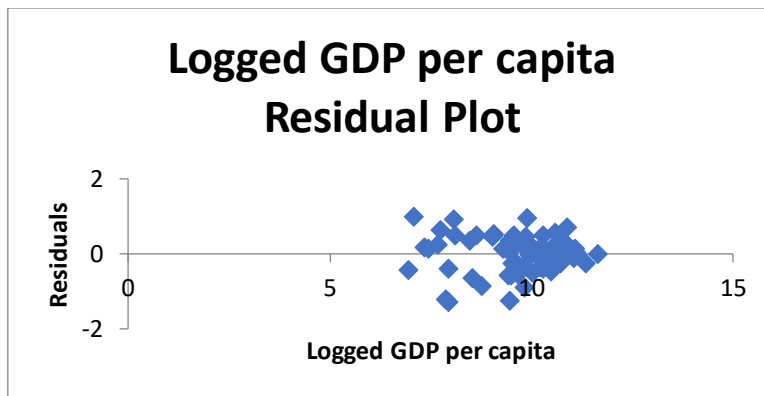|	df	SS	MS	F	Significance F
Regression	4	73.38647	18.34662	82.8076	2.29E-26
Residual	73	16.17367	0.221557		
Total	77	89.56014			

There are 4 degrees of freedom for the regression model, which corresponds to the number of independent variables (predictors) in the model.

The F-statistic is the ratio of the mean square for regression to the mean square for residuals. It assesses whether the regression model as a whole is statistically significant. In this case, the F-statistic of approximately 82.81 is very high, indicating that the regression model is statistically significant.

Finally, the very low p-value (approximately 2.29E-26) suggests that the regression model is highly significant, providing strong evidence against the null hypothesis.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-1.04567	0.733517	-1.42555	0.158259	-2.50757	0.41623
Logged GDP per capita	0.23784	0.143548	1.656875	0.101837	-0.04825	0.52393
Social support	1.978795	0.925782	2.137431	0.035912	0.133714	3.823876
Healthy life expectancy	0.061019	0.02045	2.983898	0.003869	0.020263	0.101775
Perceptions of corruption	-1.59435	0.312909	-5.09526	2.64E-06	-2.21798	-0.97072

In our analysis, we've established a significance threshold for p-values at 0.15. This means that any predictor variable with a p-value below 0.15 is considered potentially influential in explaining the variance in our dependent variable. Looking at the coefficients, we observe that all variables, including Logged GDP per capita with a p-value of 0.1018, meet this criterion. This indicates that all variables are statistically significant contributors to our model within our specified significance framework.

Logged GDP per capita Residual Plot

Social support Residual Plot

Healthy life expectancy Residual Plot

Perceptions of corruption Residual Plot

Based on these residual plots, this model demonstrates adherence to the fundamental assumption of linearity in regression analysis. The residual plots, which illustrate the differences between observed and predicted values, exhibit a consistent linear pattern. Additionally, the random distribution of residuals across the predicted values affirms that any remaining discrepancies are due to chance, reinforcing the notion of a linear

relationship between the variables. This overall alignment underscores the appropriateness of a linear regression model for your dataset.

Normal Probability Plot

R² = 0.9405
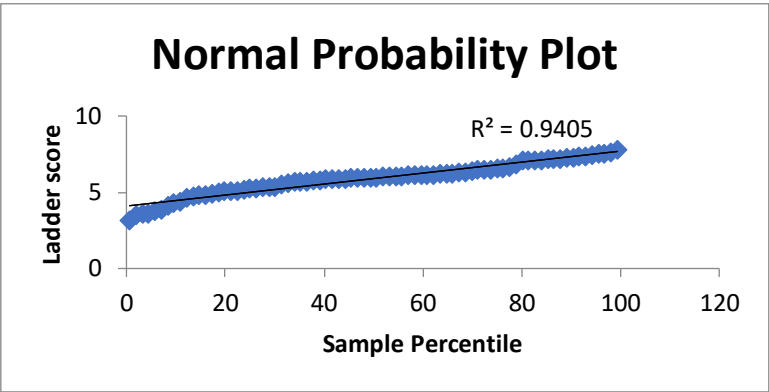
A normal probability plot is a graphical tool used to assess whether a set of data follows a normal distribution. In this case, the R-squared indicates that approximately 94.05% of the variability in the dependent variable is explained by the independent variables. This indicates that the model fits the data well and that the assumptions of normality in the residuals are met, which is a positive sign for the reliability of your regression analysis.

Appendix C: Modeling Process

In the modeling process, I started with a set of 10 variables. The initial regression analysis highlighted 'Environmental Performance Index' as having the highest p-value, indicating it should be excluded from the model.

	Coefficients	P-value
Intercept	-1.57136	0.126068
Logged GDP per capita	0.236359	0.142182
Social support	1.707455	0.083356
Healthy life expectancy	0.048355	0.027317
Freedom to make life choices	0.93023	0.329942
Generosity	-0.42791	0.346866
Perceptions of corruption	-1.41096	0.001587
Unemployment Rate	-0.01989	0.158817
Education	0.011212	0.336859
Environmental Performance Index (EPI)	0.005371	0.449971
Wealth Inequality Gini Index	0.013583	0.174213

After removing 'Environment', I observed an improvement in model performance. However, due to high correlations between 'Environment,' 'Life Expectancy,' and 'GDP,' I further evaluated which variables would enhance the model. I removed 'Life Expectancy' and 'GDP' separately, noting their impacts on the model's performance. This process helped me identify that both 'Life Expectancy' and 'GDP' were significant predictors, warranting their inclusion in the process while excluding 'Environmental Performance Index'. I then proceeded to address other variables with high p-values.

'Freedom' was removed, followed by 'Education,' leading to improved model performance. When considering 'Generosity,' I observed its high p-value and explored its impact on the model. Although removing it led to a decrease in Adjusted R Square, it also reduced the gap between R Square and Adjusted R Square, indicating an overall improvement. However, I noticed a noteworthy increase in the p-value associated with 'Unemployment.'

	Coefficients	P-value
Intercept	-1.63684	0.055743
Logged GDP per capita	0.30958	0.044462
Social support	1.90978	0.043523
Healthy life expectancy	0.055225	0.009532
Perceptions of corruption	-1.61748	4.09E-06
Unemployment Rate	-0.01525	0.229757
Wealth Inequality Gini Index	0.012992	0.173195

This prompted me to further explore the possibility of removing 'Unemployment' while retaining 'Generosity' in the model. However, after careful consideration, I determined that this adjustment ultimately led to a decrease in the Adjusted R Square and an increase in the gap between R Square and Adjusted R Square. The process continued with the elimination of both 'Unemployment' and 'Generosity', respectively. Removing them resulted in a more favorable model, with a decreased gap.
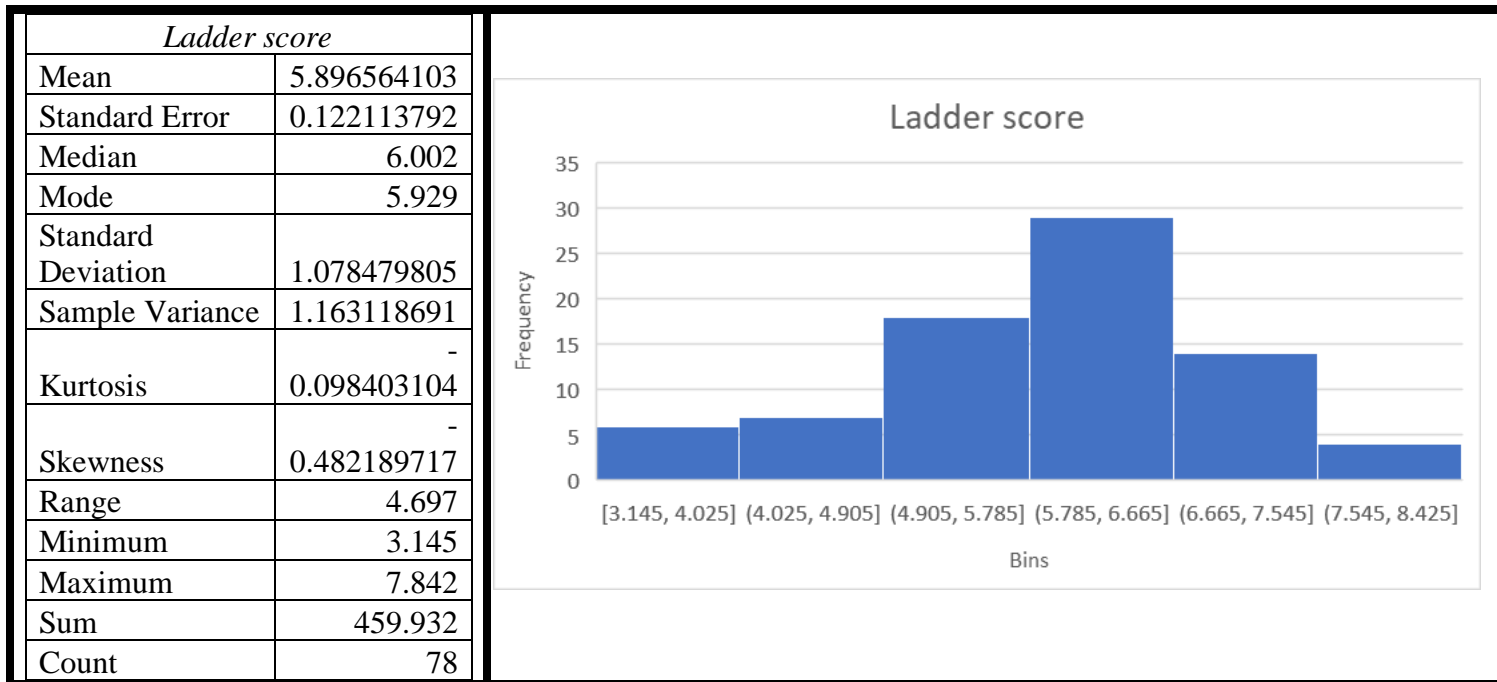
Finally, I addressed 'Inequality,' which also had a high p-value, and its removal led to a model with all p-values within an acceptable range. This final model, which exhibited the lowest gap and was considered the optimal choice, involved the following variables: Logged GDP per capita, Social Support, Healthy life expectancy, and Perceptions of corruption.

	Coefficients	P-value
Intercept	-1.04567	0.158259
Logged GDP per capita	0.23784	0.101837
Social support	1.978795	0.035912
Healthy life expectancy	0.061019	0.003869
Perceptions of corruption	-1.59435	2.64E-06

Appendix D: Data Analysis

Descriptive Statistics

Ladder Score:

Ladder score	
Mean	5.896564103
Standard Error	0.122113792
Median	6.002
Mode	5.929
Standard Deviation	1.078479805
Sample Variance	1.163118691
Kurtosis	-0.098403104
Skewness	-0.482189717
Range	4.697
Minimum	3.145
Maximum	7.842
Sum	459.932
Count	78

The average (mean) happiness score across all countries is approximately 5.90, with a standard deviation of about 1.08. This indicates that most countries tend to have happiness scores relatively close to the mean, but there is a fair amount of variability.

The scores range from a minimum of 3.145 to a maximum of 7.842, suggesting a considerable spread in happiness levels across countries.

The histogram for Ladder Score is divided into 6 bins. The distribution increases steadily up to the fourth bin, covering values between 5.785 and 6.665. After this point, there is a decrease in frequency. The distribution of Ladder Score appears to be approximately normal. This indicates that countries are evenly distributed across the range of happiness scores, with a similar number of countries experiencing both lower and higher levels of happiness.

The Ladder score exhibits strong correlations with Logged GDP per capita, Social Support, Healthy life expectancy, and EPI. For additional details, please review the correlation matrix presented in the concluding section of this appendix.

Logged GDP per Capita:

Logged GDP per capita	
Mean	9.760064103
Standard Error	0.123775185
Median	9.964
Mode	#N/A
Standard Deviation	1.093152837
Sample Variance	1.194983126
Kurtosis	-0.004820202
Skewness	-0.833174738
Range	4.689
Minimum	6.958
Maximum	11.647
Sum	761.285
Count	78
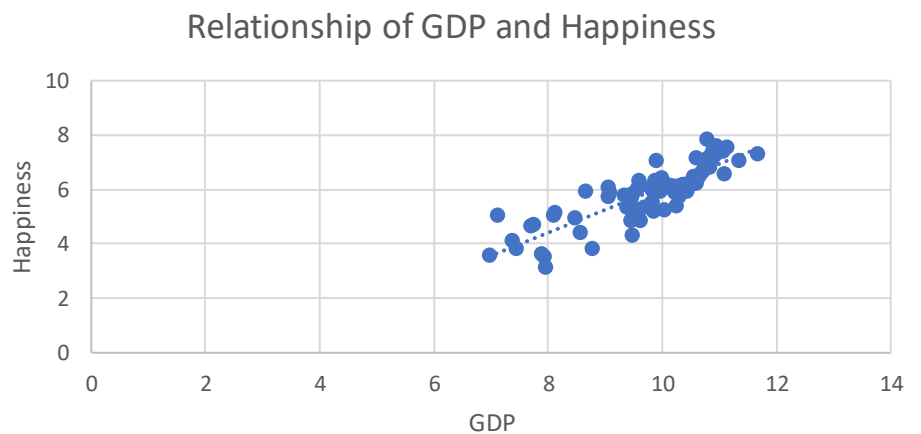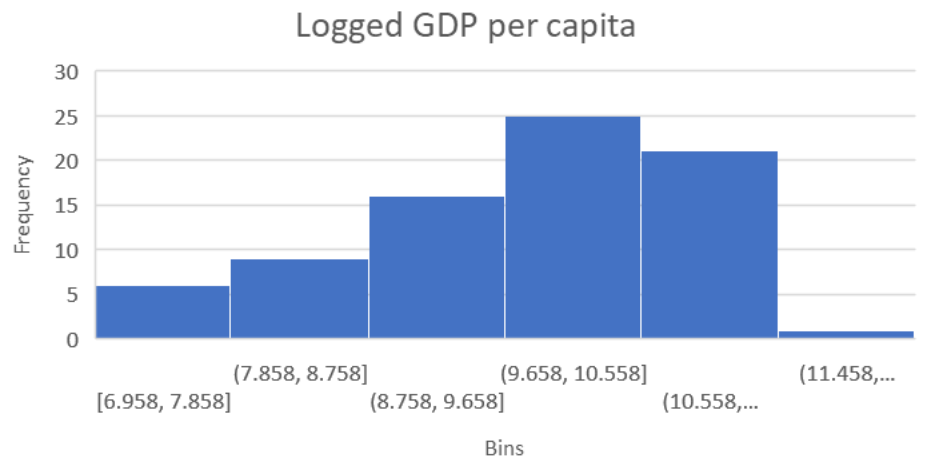
The average (mean) logged GDP per capita is roughly 9.76, with a standard deviation of about 1.09. This implies that the economic prosperity of countries in the dataset varies, with some countries having significantly higher GDP per capita than others.

The range of GDP per capita is from approximately 6.958 to 11.647. This indicates that there is a substantial disparity in economic wealth among the countries in the dataset.

The histogram for GDP is divided into 6 bins. The distribution increases steadily up to the fourth bin, covering values between 9.658 and 10.558. After this point, there is a decrease in frequency. The distribution of GDP exhibits characteristics consistent with a normal distribution, indicating a symmetrical spread of data around the mean.

A positive correlation was observed, indicating that as the GDP per capita increases, so does the Ladder Score.

GDP shows a high correlation with Social Support, Healthy life expectancy, and EPI. For additional details, please review the correlation matrix presented in the concluding section of this appendix.

Social Support:

Social support	
Mean	0.847166667
Standard Error	0.012123222
Median	0.882
Mode	0.954
Standard Deviation	0.107069393
Sample Variance	0.011463855
Kurtosis	1.669793568
Skewness	-1.39382595
Range	0.494
Minimum	0.489
Maximum	0.983
Sum	66.079
Count	78
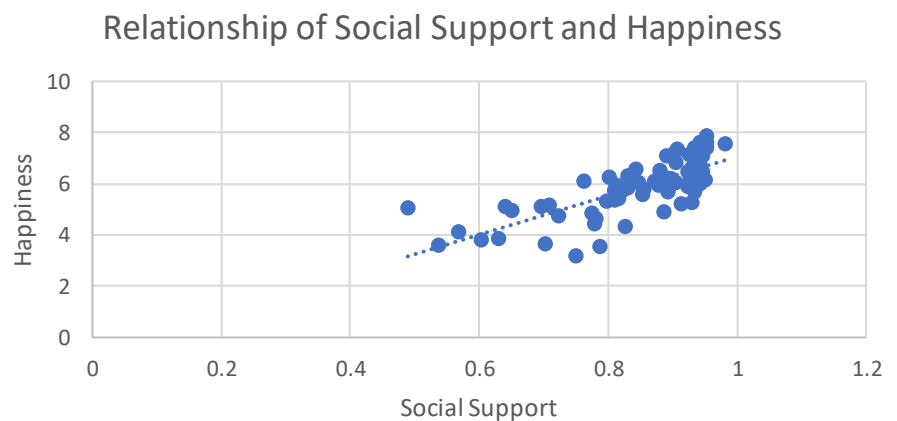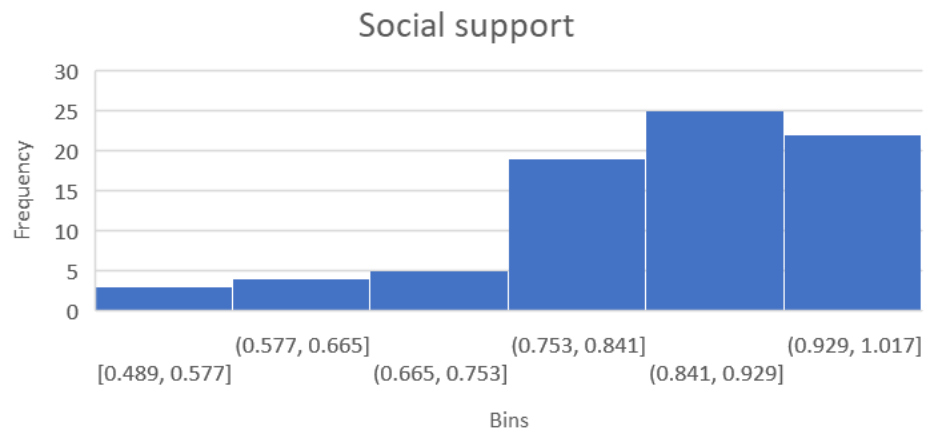
The average level of social support across all countries is approximately 0.847, with a standard deviation of about 0.107. This suggests that, on average, countries tend to have relatively high levels of social support, but there is still some variation.

The levels of social support range from a minimum of 0.489 to a maximum of 0.983, indicating that some countries have significantly higher levels of social support than others.

The histogram for Social Support also has 6 bins. The first three bins are relatively small, indicating lower frequencies. The largest bins are on the right side, particularly the middle one, covering values between 0.841 and 0.929. The last bin shows a slight decrease in frequency. The distribution of Social Support demonstrates a negative skew. This indicates that most countries tend to have relatively higher levels of social support, while fewer countries have lower levels of social support.

Social support demonstrates strong correlations with Healthy life expectancy and Ladder score. For additional details, please review the correlation matrix presented in the concluding section of this appendix.

Healthy Life Expectancy:

Healthy life expectancy	
Mean	66.97015385
Standard Error	0.7086826
Median	68.1255
Mode	72.6
Standard Deviation	6.258915252
Sample Variance	39.17402013
Kurtosis	0.399985406
Skewness	-1.052907003
Range	26
Minimum	48.7
Maximum	74.7
Sum	5223.672
Count	78
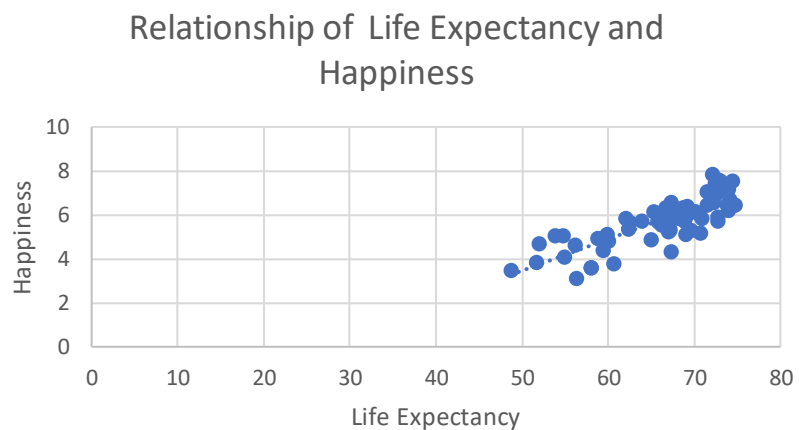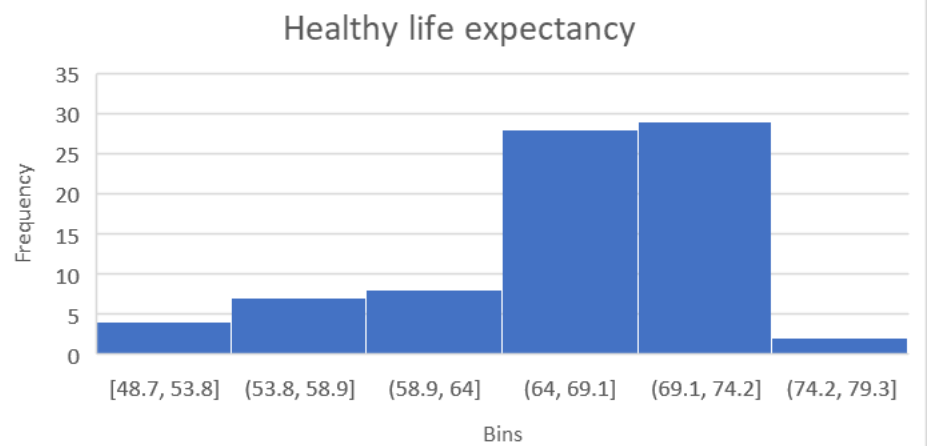
The average healthy life expectancy is approximately 66.97 years, with a standard deviation of about 6.26. This means that, on average, people in the countries tend to live around 67 years in good health, but there is a notable variation.

The range of healthy life expectancy is from approximately 48.7 to 74.7 years, showing a substantial disparity in health outcomes among the countries.

The histogram for Healthy Life Expectancy is divided into 6 bins. The largest bins are the fourth and fifth ones, representing values between 64 and 69.1, and 69.1 and 74.2 respectively. Bins 1, 2, 3, and 6 have notably lower frequencies. The distribution of Healthy Life Expectancy exhibits a negative skew. This suggests that the majority of countries have relatively higher life expectancies, while fewer countries have lower levels of healthy life expectancy.

Healthy life expectancy displays notable correlations with Ladder score, Logged GDP per capita, Social Support, and EPI. For additional details, please review the correlation matrix presented in the concluding section of this appendix.
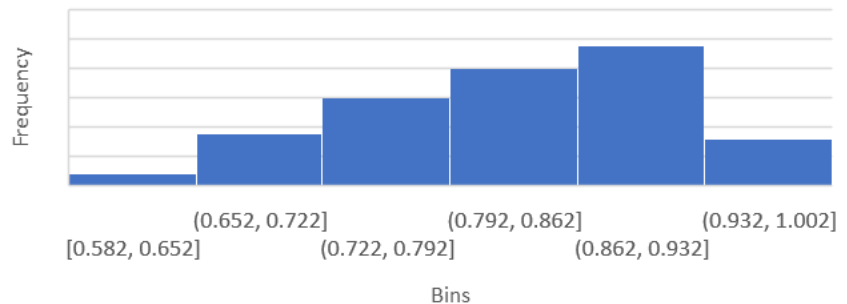
Freedom to Make Life Choices:

Freedom to make life choices	
Mean	0.829282051
Standard Error	0.00963344
Median	0.8435
Mode	0.822
Standard Deviation	0.085080237
Sample Variance	0.007238647
Kurtosis	-0.2045246
Skewness	-0.610288784
Range	0.378
Minimum	0.582
Maximum	0.96
Sum	64.684
Count	78
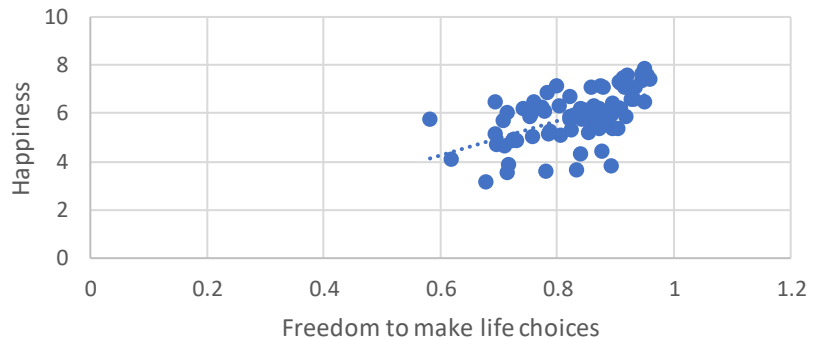
The average level of freedom to make life choices is about 0.829, with a standard deviation of around 0.085. This indicates that, on average, countries provide a relatively high degree of freedom to their citizens, but there is still some variability.

The levels of freedom range from a minimum of 0.582 to a maximum of 0.96, suggesting that some countries offer significantly more freedom than others.

The histogram for Freedom to Make Life Choices is divided into 6 bins. The frequency increases steadily up to the fifth bin, covering values between 0.862 and 0.932. After this point, there is a decrease in frequency. The distribution of Freedom to Make Life Choices is negatively skewed. This implies that a larger number of countries have relatively higher levels of freedom, while fewer countries experience lower levels of freedom.

A moderate positive correlation was noted, implying that greater freedom leads to higher levels of reported happiness.

Generosity:

Generosity	
Mean	-0.017102564
Standard Error	0.016838923
Median	-0.04
Mode	-0.034
Standard Deviation	0.148717338
Sample Variance	0.022116846
Kurtosis	2.988522775
Skewness	1.270144033
Range	0.83
Minimum	-0.288
Maximum	0.542
Sum	-1.334
Count	78
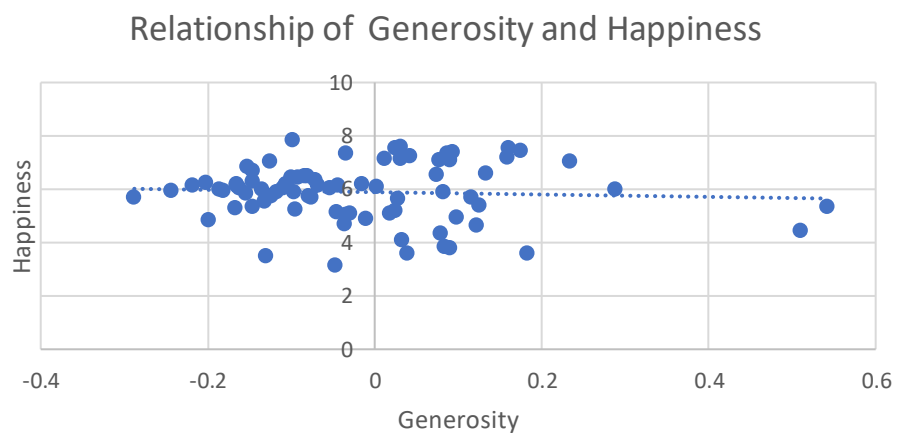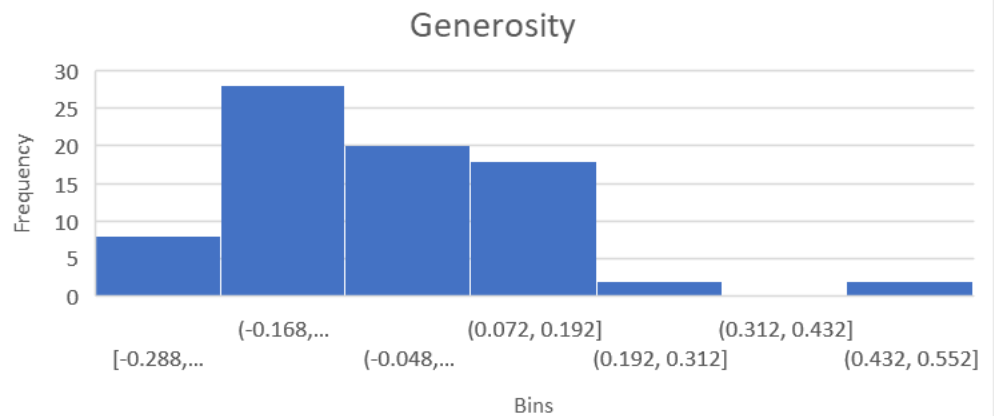
The average level of generosity is approximately -0.017, with a standard deviation of about 0.149. This suggests that, on average, countries tend to have a slightly negative generosity score, but there is still some variation.

The generosity scores range from a minimum of -0.288 to a maximum of 0.542, indicating that there is a range of attitudes towards generosity across countries.

The histogram for Generosity comprises 7 bins. The first bin is small, followed by the largest bin, covering values between -0.168 and -0.048. Frequency then decreases, and the sixth bin, with values between 0.312 and 0.432, has no data. The Generosity distribution appears to be positively skewed. This indicates that there are more countries with lower levels of reported generosity, and fewer countries with very high levels of generosity.

While the trend line appears relatively straight, it shows a negative correlation. This suggests that higher levels of reported generosity may not necessarily lead to higher Ladder Scores.

Perceptions of Corruption:

Perceptions of corruption	
Mean	0.71624359
Standard Error	0.021975997
Median	0.794
Mode	0.809
Standard Deviation	0.194086752
Sample Variance	0.037669667
Kurtosis	0.710840488
Skewness	-1.256447584
Range	0.76
Minimum	0.179
Maximum	0.939
Sum	55.867
Count	78

The average perception of corruption is around 0.716, with a standard deviation of about 0.194. This means that, on average, countries tend to have a relatively low perception of corruption, but there is still some variation.

The perceptions of corruption range from a minimum of 0.179 to a maximum of 0.939, indicating that some countries are perceived to have higher levels of corruption than others.

The histogram for Perception of Corruption shows an increasing trend up to the last bin, which is the largest. This bin covers values between 0.819 and 0.979. The distribution of Perception of Corruption shows a negative skew. This suggests that most countries tend to have relatively lower levels of perceived corruption, while fewer countries have very high levels of perceived corruption.

A negative correlation was evident, indicating that countries with lower levels of perceived corruption tend to have higher Ladder Scores.

Unemployment Rate:

Unemployment Rate	
Mean	7.18974359
Standard Error	0.51417915
Median	6.03
Mode	5.17
Standard Deviation	4.5411069
Sample Variance	20.6216151
Kurtosis	3.9346704 25
Skewness	1.8344886 51
Range	23.85
Minimum	0.75
Maximum	24.6
Sum	560.8
Count	78
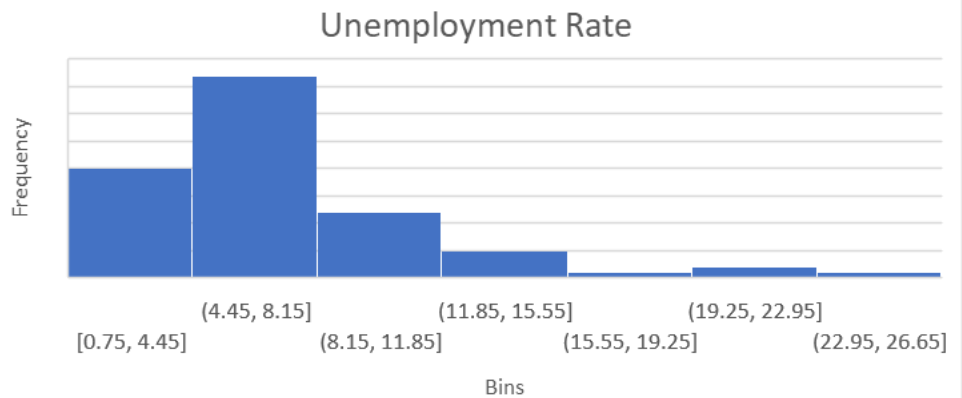
The average unemployment rate in 2021 is approximately 7.19%, with a standard deviation of about 4.54. This suggests that, on average, countries tend to have a relatively moderate level of unemployment, but there is still some variation.

The unemployment rates range from a minimum of 0.75% to a maximum of 24.6%, indicating a significant disparity in employment situations among the countries.

The histogram for Unemployment is divided into 7 bins. The second bin is the largest, representing values between 4.45 and 8.15. After this point, there is a decrease in frequency, followed by a slight increase and another decrease. The distribution of Unemployment displays a slight positive skew. This indicates that most countries tend to have lower unemployment rates, while fewer countries experience higher rates of unemployment.

Although the trend line appears relatively straight, it demonstrates a negative correlation. This implies that lower unemployment rates are associated with higher reported happiness levels.

Education:

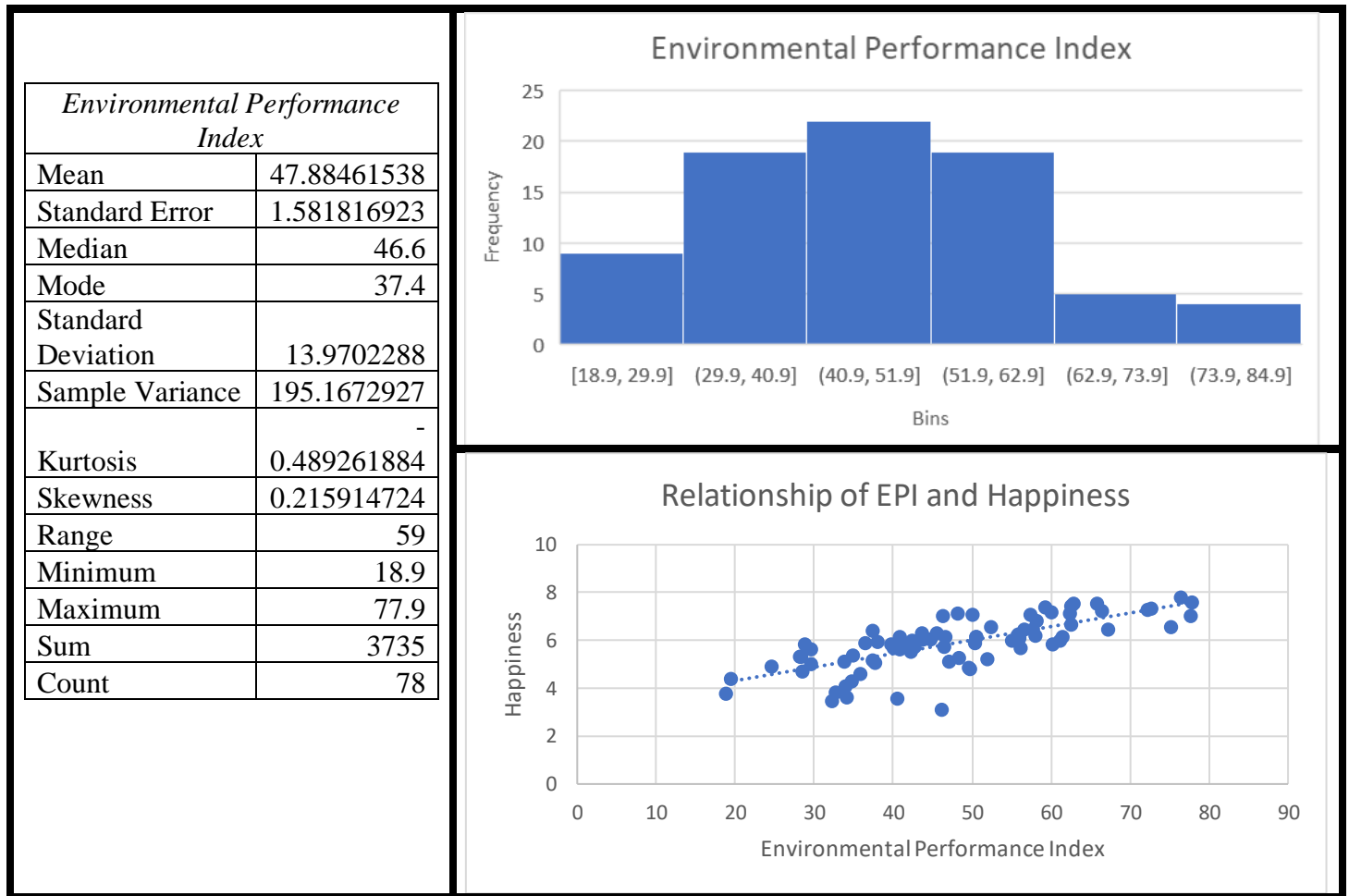| Education	
Mean	9.722820513
Standard Error	0.789146803
Median	9.17
Mode	#N/A
Standard Deviation	6.969555856
Sample Variance	48.57470882
Kurtosis	-0.860110043
Skewness	0.359478174
Range	26.65
Minimum	0.15
Maximum	26.8
Sum	758.38
Count	78
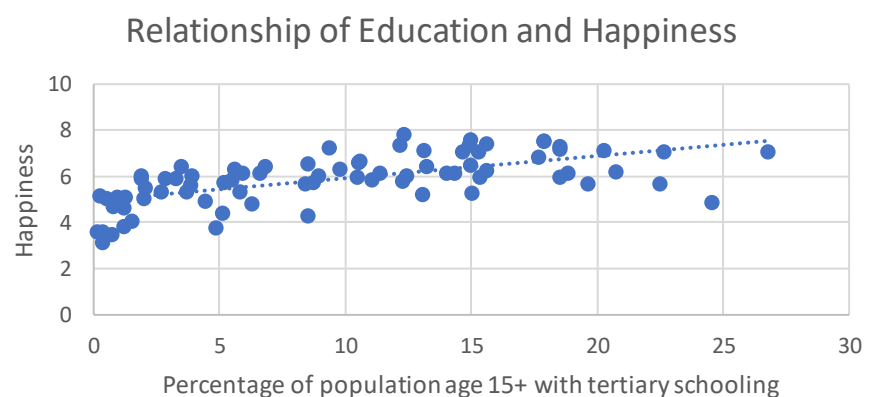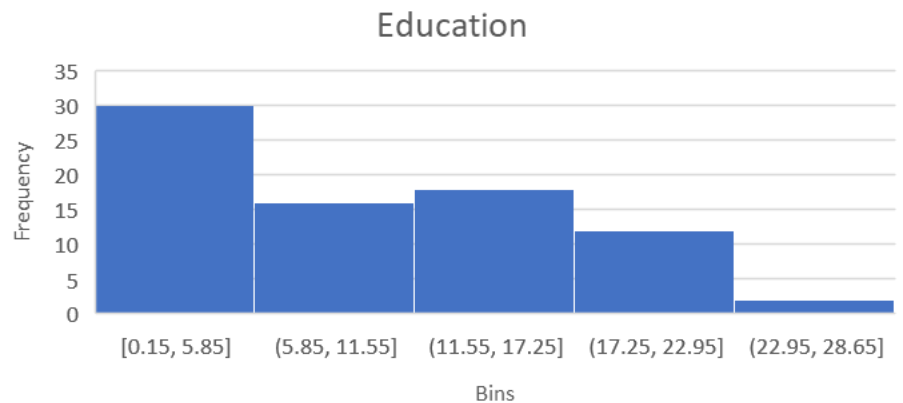
The average percentage of the population (age 15+) with completed tertiary schooling in 2010 is approximately 9.72%. The standard deviation is 6.97, indicating some variability in educational attainment across countries.

The percentages range from a minimum of 0.15% to a maximum of 26.8%, showing a considerable range in educational attainment levels.

The histogram for Education is divided into 5 bins. The first bin is the largest, covering values between 0.15 and 5.85. Frequency then decreases, followed by a slight increase and another decrease. The distribution of Education appears to be positively skewed. This implies that a larger number of countries have lower percentages of the population with completed tertiary schooling, while fewer countries have very high percentages.

A positive correlation was identified, signifying that countries with a higher percentage of the population completing tertiary schooling tend to have higher Ladder Scores.

Environmental Performance Index (EPI):

Environmental Performance Index	
Mean	47.88461538
Standard Error	1.581816923
Median	46.6
Mode	37.4
Standard Deviation	13.9702288
Sample Variance	195.1672927
Kurtosis	-0.489261884
Skewness	0.215914724
Range	59
Minimum	18.9
Maximum	77.9
Sum	3735
Count	78

Environmental Performance Index histogram with bins [18.9, 29.9] (29.9, 40.9] (40.9, 51.9] (51.9, 62.9] (62.9, 73.9] (73.9, 84.9]

Relationship of EPI and Happiness scatterplot

The average Environmental Performance Index (EPI) is about 47.88, with a standard deviation of approximately 13.97. This suggests that, on average, countries tend to have a moderate EPI score, but there is still a considerable amount of variability.

The EPI scores range from a minimum of 18.9 to a maximum of 77.9, indicating a wide range of environmental performance among the countries.

The histogram for EPI is divided into 6 bins. The distribution increases up to the third bin, which is the largest, covering values between 40.9 and 51.9. After this point, there is a decrease in frequency. The EPI distribution appears to be approximately normal. This indicates that countries are evenly distributed across the range of EPI scores, with a similar number of countries experiencing both lower and higher levels of environmental performance.

Environmental Performance Index (EPI) shows significant correlations with Ladder score, Logged GDP per capita, and Healthy life expectancy. For additional details, please review the correlation matrix presented in the concluding section of this appendix.

Wealth Inequality Gini Index:

Wealth Inequality Gini Index	
Mean	35.69358974
Standard Error	0.78399818
Median	35.15
Mode	26
Standard Deviation	6.924084444
Sample Variance	47.94294539
Kurtosis	-0.292931679
Skewness	0.511771735
Range	28.9
Minimum	24
Maximum	52.9
Sum	2784.1
Count	78
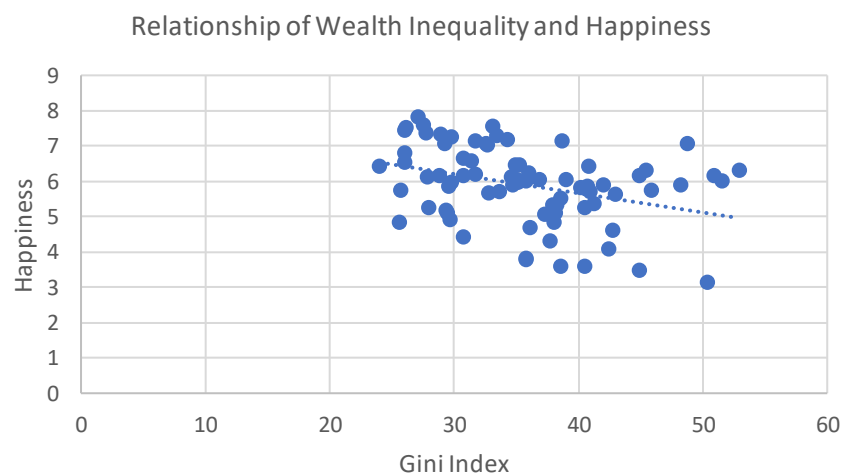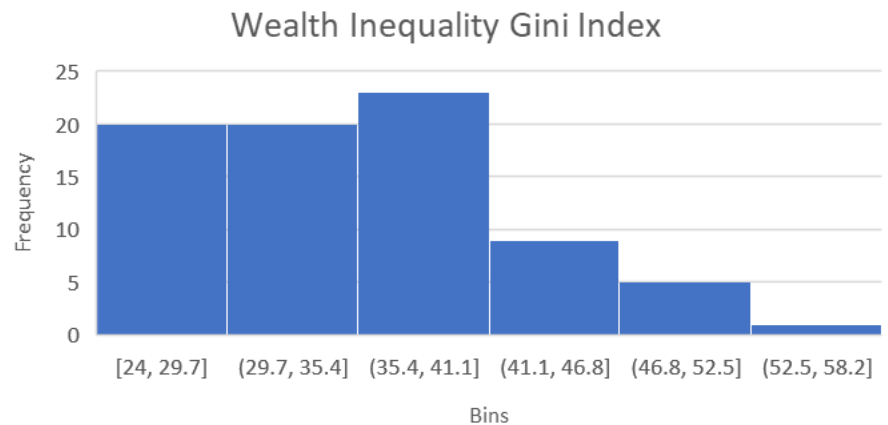
The average Gini Index for wealth inequality is roughly 35.69, with a standard deviation of about 6.92. This means that, on average, countries tend to have a moderate level of wealth inequality, but there is still some variation.

The Gini Index for wealth inequality ranges from a minimum of 26 to a maximum of 52.9, showing that there are differences in wealth distribution among the countries.

 The histogram for Wealth Inequality is divided into 6 bins. The first two bins are equal in frequency. The third bin is the largest, covering values between 35.4 and 41.1. After this point, there is a decrease in frequency. The distribution of Wealth Inequality shows a positive skew. This implies that the majority of countries tend to have relatively lower wealth inequality (closer to equality), while fewer countries exhibit very high levels of wealth inequality.

A negative correlation was found, indicating that countries with lower wealth inequality tend to have higher Ladder Scores.

Correlations

	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
Social support	1				
Healthy life expectancy	0.766337	1			
Freedom to make life choices	0.434816	0.486899	1		
Generosity	-0.09794	-0.14439	0.242908	1	
Perceptions of corruption	-0.30434	-0.38305	-0.54219	-0.24767	1
Unemployment Rate 2021	0.079925	0.018814	-0.2343	-0.42868	0.183577
Education	0.615204	0.623295	0.177633	-0.06042	-0.3575913
Environmental Performance Index (EPI)	0.639403	0.705089	0.307607	-0.18231	-0.5553965
Wealth Inequality Gini Index	-0.29237	-0.3297	-0.21293	-0.15853	0.4022288

	Ladder score	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices
Ladder score	1				
Logged GDP per capita	0.855398262	1			
Social support	0.756965006	0.8371414	1		
Healthy life expectancy	0.833359007	0.9075263	0.7663	1	
Freedom to make life choices	0.562364237	0.5007471	0.4348	0.48689897	1
Generosity	-0.060415302	-0.125294	-0.0979	-0.14439384	0.24290837
Perceptions of corruption	-0.590314676	-0.447815	-0.3043	-0.38304647	-0.542192716
Unemployment Rate 2021	-0.065259451	0.0275667	0.0799	0.018814472	-0.234295056
Education	0.622812039	0.6857911	0.6152	0.623294549	0.177632734
Environmental Performance Index (EPI)	0.732195057	0.7506702	0.6394	0.705088582	0.307606968
Wealth Inequality Gini Index	-0.349681373	-0.438279	-0.2924	-0.32970326	-0.21293353

	Unemployment Rate	Education	Environmental Performance Index (EPI)	Wealth Inequality
Unemployment Rate 2021	1			
Education	0.079882	1		
Environmental Performance Index (EPI)	0.031507	0.585327	1	
Wealth Inequality Gini Index	0.270576	-0.39649	-0.43127	1

Strong Correlations:

Countries with higher GDP tend to have higher ladder scores, indicating a strong relationship between economic prosperity and happiness.

Higher social support is highly associated with higher ladder scores, indicating the importance of social connections for well-being.

Longer life expectancies are positively correlated with higher ladder scores, reflecting the significance of health for happiness.

Better environmental performance is associated with higher ladder scores, suggesting a link between environmental quality and happiness.

Countries with higher GDP per capita tend to have better environmental performance, indicating a strong positive relationship between economic prosperity and environmental quality.

Healthy Life Expectancy and Environmental Performance Index also have a strong positive correlation, indicating that countries with higher environmental performance tend to have longer healthy life expectancies.

There is a strong positive correlation between Social Support and Logged GDP per capita. This indicates that countries with higher GDP per capita tend to have higher levels of social support.

Healthy Life Expectancy and Logged GDP per capita are highly correlated. This suggests that countries with higher economic prosperity tend to have longer healthy life expectancies.

There is a strong positive correlation between Healthy Life Expectancy and Social Support. This indicates that countries with higher levels of social support tend to have longer healthy life expectancies.

Appendix E: Data

As previously mentioned, I gathered data from various sources, and each offered insights from different years. For the education dataset, I specifically focused on the most recent data available in 2010. Similarly, for the 'World Happiness Reports' dataset, I relied on data from the year 2021. When it came to the unemployment dataset, I had access to rates spanning multiple years, but I opted for the latest available data from 2021. In the case of the 'Wealth Inequality' dataset, there were data points from different years for each country. My approach prioritized the use of the most current data available for each source. I merged the datasets based on countries. This allowed for a comprehensive analysis that takes into account the latest available data across various facets of national well-being.

It's noteworthy that outliers were detected in the variables Generosity, Unemployment, EPI, and Education. Despite their presence, I decided to keep them in the analysis, given that our focus is on country-level data. Consequently, no countries were excluded from

the study based on these outliers because they are part of the market we are interested in.