

## A Real Data-Driven Analytical Model to Predict Happiness

Aditya Chakraborty<sup>1\*</sup>, Dr. Chris P. Tsokos<sup>2</sup>

<sup>1</sup>Doctoral Candidate, Department of Mathematics & Statistics, University of South Florida

<sup>2</sup>Distinguished University Professor Department of Mathematics & Statistics, University of South Florida

DOI: [10.36347/sjpm.2021.v08i03.001](https://doi.org/10.36347/sjpm.2021.v08i03.001)

| Received: 21.02.2021 | Accepted: 03.03.2021 | Published: 09.03.2021

\*Corresponding author: Aditya Chakraborty

"This research is Copyright protected by University of South Florida with ID: 20B127"

### Abstract

### Original Research Article

**Purpose:** Philosophers and many modern-day researchers are convinced by the fact that the pursuit of happiness is the ultimate goal for humankind. Aristotle believed that the utmost goal of human life was eudaimonia (interpreted as "happiness," "human flourishing," or "a good life."). Recently, many economists and physiologists have been doing applied research in the areas of subjective well-being (SWB) or happiness and trying to understand how it improves the quality of life of individual beings. Thus, searching for a data-driven analytical model is crucial to predict SWB and enhance the quality of life. **Methods:** Our present study utilizes the world happiness database obtained from the GallupWorld Poll on the happiness of 156 countries. However, our study focuses on using only the data of fiftyfour developed countries, based on the *human development index (HDI)*. We have developed a non-linear analytical model that predicts the average happiness score based on eleven risk factors with a high degree of accuracy. We also compared our analytical model with three other statistical models, and our model outperformed the rest of the three in terms of RMSE and MAE. **Results:** Our analytical model includes five important findings. The response of the proposed model is the average score of happiness of individuals in developed countries. In addition to predicting the happiness score, our model identifies the individual risk factors and their corresponding interactions that significantly contribute to happiness. We rank these risk factors by their percentage of contributions to the happiness score. We also proceed to rank the developed countries with respect to their predicted happiness score from our developed model. From our study, we found Finland being number one, followed by Denmark. The U.S is fifth and Romania being 54th. **Conclusion:** The proposed model offers other useful information on the subject area. Our analytical model has been validated and tested to be of high quality, and our prediction of happiness is with a high degree of accuracy. We created a survey questionnaire (appendix 1) based on the data that can be used along with our model by any company for the strategic planning or decision making. **Keywords:** Gallup world poll, Subjective Well Being (SWB), nonlinear statistical modelling, Machine learning regularization techniques.

Copyright © 2021 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC 4.0) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

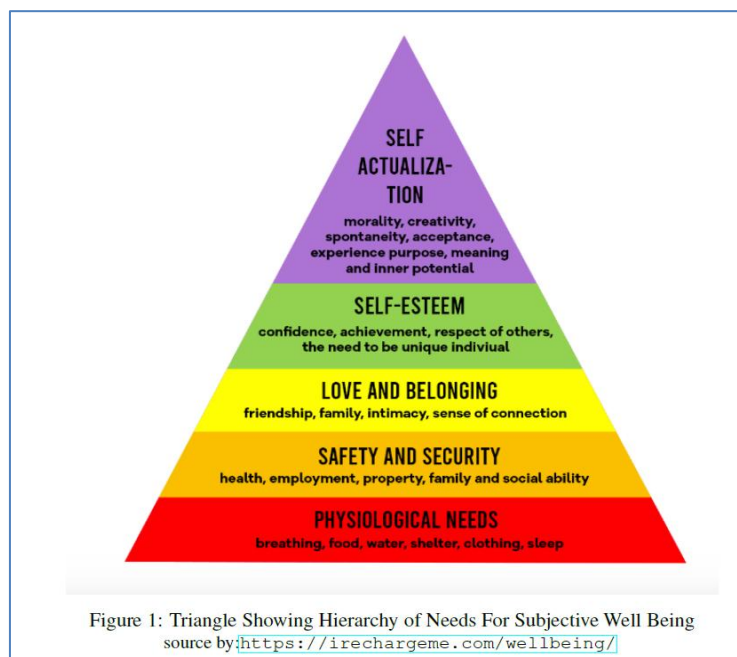
## INTRODUCTION

Aristotle believed that the ultimate aim of human life was a concept of ancient Greeks called eudaimonia. The word is often translated as Happiness, but more likely means "human flourishing" or "a good life." Rather than an emotion or temporary mood, eudaimonia is better evaluated by asking the question, "What do I want to be remembered for when I will be no more on earth? Being happy is not only associated with personal well-being but also with the productivity on a large scale. Studies have been performed to understand the association between happiness and

productivity [1]. Happy individuals tend to perform better and lower Happiness is correlated with lower productivity. These different forms of evidence, with complementary strengths and weaknesses, are consistent with the existence of a causal link between human well-being and human performance [1]. A happy mind is also associated with sound mental health. Health and Happiness are essential and possibly related to the pursuits of humankind. Health may be a determinant of Happiness or, at least, ill-health may cause unhappiness. Conversely, a feeling of Happiness may enhance health [2]. Most research on Happiness has been done by social scientists are focused on

psychological and social determinants of Happiness. For instance, Happiness is routinely monitored in sociologic surveys [6], and levels of Happiness have been associated with personality traits [3], living conditions [2], self-esteem [7], feeling in love [4], democracy [5], but also with specific loci of brain activity [8]. Few studies have examined Happiness concerning health in a general population. An epidemiologic study of Finnish men indicated that life satisfaction (measured through four items assessing whether life is interesting, happy, easy, or lonely) predicts lower mortality [9], but the specific contribution of Happiness was not reported. This is rather typical of the medical literature, where Happiness is often considered to be but one component of good mental health. For instance, the mental health scale embedded in the Short Form-36 (SF-36) questionnaire includes an item on happiness [10], one item from the Bradburn the scale of well-being asks whether the respondent is 'depressed or very unhappy' [11, 12], and the validity of the Happiness-Depression scale was tested against a mental health questionnaire [13]. Taking the opposite stand, Saracci has argued that the World Health Organization definition of health would be more appropriate for a definition of Happiness, and

that health requires a narrower definition [14]. However, how the concepts of health, well-being, quality of life, and Happiness relate to each other remains a matter of debate. Some studies also found that the effect of the nationality of levels of Happiness may capture the impact of cultural integration on people's well-being [2]. Using an international cross-section of 28 countries, researchers have found a highly significant impact of democracy on the subjective well-being of people [19]. Thus, Happiness and democracy, as one would expect, are highly correlated. In general, personal Happiness and well-being seem to be the principal objective of human life. Throughout history, the virtue of Happiness has been considered as the ultimate end of temporal existence. Aristotle's ancient view about Happiness was "Happiness is so important; it transcends all other temporal considerations." Aristotle's prescription for spending a good life was to exercise virtues like to be kind, humble, wise, and honest in our actions consistently. In other words, accomplishing different physical and emotional needs, as listed in Figure 1, is the recipe for a happy life. The following Figure 1 illustrates briefly different stages of human needs to achieve Subjective Well-Being.



While building the statistical model, the response variable is the national average of happiness Score; hence, we develop an analytical model containing significant contributable variables and other significant interactions. The proposed statistical model relies on several assumptions, such as linearity, multicollinearity, homoscedasticity, and different assumptions concerning statistical methodology. The dataset shows that some of the risk factors are highly correlated, as shown in Figure 4; thus, the parameters of the models become difficult to interpret. The parameters

also become very unstable when independent variables are highly correlated, which leads to over-fitting the model. Moreover, we use different penalization regression methods: Ridge Regression ( $L_2$ ) [14], Lasso Regression ( $L_1$ ) [15], and Elastic net (EN) [16]. These methods are vastly used to address the over-fitting of the model. Our proposed statistical model is useful in predicting individuals' Happiness, given the values of the significant risk factors. Also, we ranked the risk factors according to their percentage of contribution to the happiness score. The validation and quality of our

proposed analytical model have been statistically evaluated using R Square ( $R^2$ ), R Square Adjusted ( $R_{adj}^2$ ), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and residual analysis. Eventually, its usefulness has been illustrated by utilizing different combinations of various risk factors. To the best of our knowledge, no such statistical model has been developed under the proposed logical structure to predict Happiness for developing countries. Therefore, searching for an appropriate statistical model in the prediction of Happiness is imperative.

## METHODOLOGY

### The Data

The World Happiness Report is a landmark survey of the state of global happiness that ranks descriptively 156 countries by how happy their citizens perceive themselves to be. The data has been obtained from the World Happiness Report 2019 website, where they used the Gallup Poll to get the answers to specific questions (appendix 1). The data has been collected for a total of 156 countries from 2005 to 2018. However, in our study, we only considered the data of developed countries (sorted based on the human development index [HDI]) in the world. Individuals were asked specific questions, and as a result of their response as a whole, a score was produced, which is termed as the national average. In our data, the average scores of the developed countries from 2005 to 2018 were tabulated. One of the main goals of our study is to understand what attributable variables significantly affect the happiness of an individual. We have eleven attributable variables and the response Ladder (which is also called subjective well-being [SWB] or happiness score) as a measure of response. For example, let there is an imaginary ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life, and the bottom of the ladder represents the worst possible life of an individual. On which step of the ladder is an individual standing currently is reported. This measure is also referred to as the Cantril life ladder or just life ladder in our analysis.

The description of the attributable variables (risk factors) that the data was collected on are given below.

**LOG\_GDP( $X_1$ ):** Per-capita gross domestic product (in logarithmic scale) in purchasing power parity(PPP).

**SOC\_SUPPORT( $X_2$ ):** (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?”

**LIFE\_EXPECT( $X_3$ ):** Healthy life expectancies at birth are based on the data extracted from the World Health Organization’s (WHO) Global Health Observatory data repository.

**FREEDOM( $X_4$ ):** 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( $X_5$ ):** 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.

**PER\_CORR( $X_6$ ):** 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.

**POS\_AFFECT( $X_7$ ):** Positive affect is defined as the average of three positive affect measures in GWP: happiness, laugh and enjoyment in the Gallup World Poll.

**NEG\_AFFECT( $X_8$ ):** Negative affect is defined as the average of three negative affect measures in GWP. They are worry, sadness and anger, respectively.

**CONF\_GOV( $X_9$ ):** How much trust and confidence do you have in government when it comes to handling [International problems/Domestic problems] – a great deal, a fair amount, not very much or none at all?

**DEM\_QUALITY( $X_{10}$ ):** National average of the first two dimensions of World Governance Index (WGI) namely, voice and Accountability and Political Stability and Absence of Violence/Terrorism.

**DEL\_QUALITY( $X_{11}$ ):** National average of the last two dimensions of World Governance Index (WGI) namely, Government Effectiveness, Regulatory Quality, Rule of Law and Control of Corruption.

The definitions of the above mentioned measures under DEM\_QUALITY and DEL\_QUALITY (which are also the six dimensions of the World Governance Quality Index (WGI) are as follows:

1. Voice and Accountability: Voice and accountability captures perceptions of the extent to which a country’s citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media.
2. Political Stability and Absence of Violence/Terrorism: Political Stability and Absence of Violence/Terrorism measures perceptions of the likelihood of political instability and/or politically motivated violence, including terrorism.
3. Government Effectiveness: Government effectiveness captures perceptions of the quality of public services, the quality of the civil service on the degree of its independence from political pressures, the quality of policy

formulation and implementation, and the credibility of the government’s commitment to such policies.

4. Regulatory Quality: Regulatory quality captures perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.
5. Rule of law: Rule of law captures perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.
6. Control of corruption: Control of corruption captures perceptions of the extent to which

public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests.

From Figure 2 below, we see that there are some missing observations in the data set. However, the proportion of missing values is small; we used the predictive mean matching (PMM) algorithm [20] to perform multiple imputations to our dataset. Predictive mean matching (PMM) is a semi-parametric technique to perform multiple imputations [21] for missing data points in a plausible manner, especially for imputing quantitative variables that are not normally distributed.

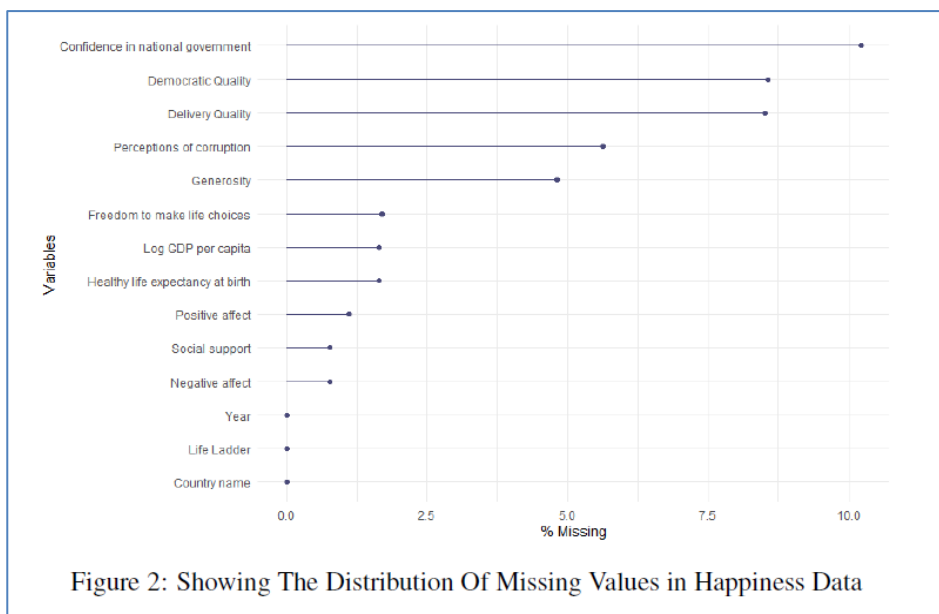


Figure 2: Showing The Distribution Of Missing Values in Happiness Data

In developing the proposed statistical model for happiness score as a function of the attributable variables, one of the main assumptions is the normality of response. That is, the response variable Ladder

should follow the Gaussian probability distribution. The mid-values of happiness score seem to be reasonably straight, but the ends are somewhat skewed, as can be seen from the Q-Q plot shown in Figure 3.

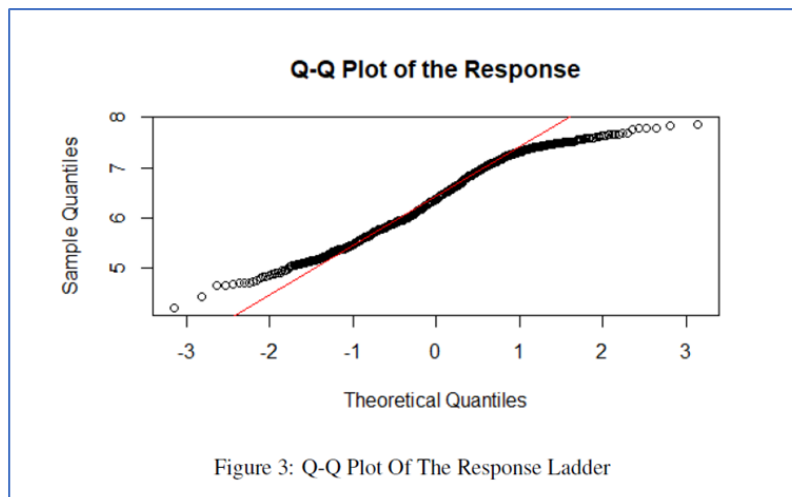


Figure 3: Q-Q Plot Of The Response Ladder

We have shown through goodness-of-fit testing (Shapiro-Wilk normality test,  $p\text{-value} = 5.565 \times 10^{-10}$ ) that the response Ladder does not follow the normal probability distribution as well. Therefore, the Q-Q plot supports the fact that natural phenomena, such as the national average of happiness scores, are not following the Gaussian probability distribution. The collinearity assumption of the model is shown in Figure 4, where negative correlations are displayed in red and positive correlations in blue color. The color intensity and the degree of the relationship between each pair are proportional to the correlation coefficients. From the following correlation matrix in Figure 4, we see that

there are strong positive associations between the variables LIFE\_EXPECT and DEL\_QUALITY, Generosity, and DEL\_QUALITY and DEM\_QUALITY and DEL\_QUALITY. Also, there is a strong negative association between the variables LOG\_GDP and PER\_CORR and PER\_CORR and CONF\_GOV. Thus, at this point, we would consider regularization techniques such as Ridge Regression ( $L_1$ ), Lasso Regression ( $L_2$ ), and Elastic net penalties to address over-fitting. Lastly, there are statistically significant relationships between response Ladder and the attributable variables (risk factors) to our nonlinear regression model.

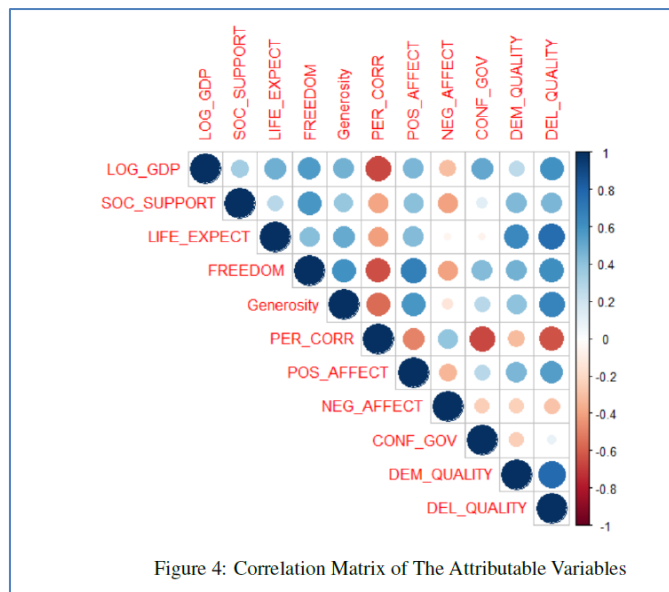


Figure 4: Correlation Matrix of The Attributable Variables

**Development of Statistical Model**

We now proceed to develop the statistical model, which is driven by the national average of happiness score as a function of the eleven risk factors and all possible interactions, as previously discussed.

One of the pure forms of a model with all possible interactions and additive error structure, in the present situation, could be expressed as follows:

$$Ladder = \beta_0 + \sum_i \alpha_i x_i + \sum_j \gamma_j k_j + \epsilon \tag{1}$$

where  $\beta_0$  is the intercept of the model,  $\alpha_i$  is the coefficient of  $i^{th}$  individual attributable variable  $x_i$ ,  $\gamma_j$  is the coefficient of  $j^{th}$  interaction term  $k_j$  and  $\epsilon$  denotes the random disturbance or residual error of the model that follows a normal distribution with zero mean and constant variance.

follow the Gaussian probability distribution. As we illustrated above (from Figure 3), the dependent variable Ladder does not follow the Gaussian probability distribution initially. Therefore, we proceed with a non-linear transformation to the happiness data to filter our data to follow the normal probability distribution. We used Johnson transformation for our response, which results in equation 2, below:

One of the underlying assumptions to develop the above model is that the response variable should

$$z = \gamma + \delta \ln \left[ \frac{(x - \epsilon)}{(\lambda + \epsilon - x)} \right], \quad \epsilon < x < \epsilon + \lambda,$$

and



$$TLadder = -0.43 + 0.87 \ln \left[ \frac{(x - 4.2)}{(3.72 + 4.2 - x)} \right] \tag{2}$$

Here, *T Ladder* denotes the new response variable (transformed) after the use of Johnson  $S_U$  Trans-formation to our old response. We now estimate the coefficients (weights) of the risk factors for the processed data in equation 2. To develop our analytical model, we initially proceed with the full statistical model, including all eleven risk factors and ten possible interactions between each pair. Thus, initially, we start structuring our model with  $n_{C_k} = 55$  ( $n = 11$ ;  $k = 2$ ) terms that include the primary contribution of the attributable variables and all possible interactions. Since we started with the full statistical model (fifty-five terms), as mentioned above, we have applied the backward elimination process to determine the most significant contributions of both the individual attributable variables and interactions by eliminating the less important risk factors gradually. Moreover,

backward elimination is considered one of the best traditional methods for a small set of feature vectors to tackle the problem of overfitting and perform feature selection.

However, the statistical estimation process of our data analysis has shown that only seven out of the eleven risk factors significantly contribute and twenty-eight interaction terms. Thus, the best proposed statistical model with all significant risk factors and interactions that estimates the happiness score accurately are seven out of eleven risk factor individually that significantly contribute and the twenty-eight interaction term. Thus, the best proposed statistical model with all significant risk factors and interactions that estimates the national average happiness score is given by equation (3) below.

$$TLadder = \begin{cases} -.45 + .42X_1 + .12 \exp(X_2) + .04X_3 + .27X_4 + .16X_5 \\ +.03 \exp(-X_6) + .13 \exp(X_7) - .07X_8 + .04X_9 \\ -.07X_{10} - .11X_{11} - .17X_1X_3 + .14X_1X_4 - .13X_1X_5 \\ +.27X_1X_6 - .03X_1X_7 - .11X_1X_8 + .22X_1X_{11} - .1X_2X_5 \\ +.19X_2X_6 + .14X_2X_8 + .22X_2X_9 + .16X_2X_{10} + .15X_2X_{11} \\ +.07X_3X_7 - .06X_3X_{10} - .43X_4X_6 - .19X_4X_8 - .29X_4X_9 \\ +.1X_4X_{10} - .3X_4X_{11} + .18X_5X_6 + .11X_5X_9 - .2X_5X_{10} \\ +.32X_5X_{11} - .02X_6X_{11} + .1X_7X_8 + .1X_7X_9 + .1X_8X_{11} \end{cases} \tag{3}$$

$TLadder$  is obtained from equation (3) and is based on the Johnson transformation(18) of the data, thus, we will utilize the anti-transformation on equation

(3) to estimate the desired, actual national average of happiness score as follows:

$$Ladder = 4.2 + \frac{3.72}{1 + \exp\left(\frac{TLadder + .43}{.87}\right)} \tag{4}$$

The proposed model will help social scientists and economists understand how the happiness score changes when any of the eleven attributable variables is varied by keeping the other risk factors fixed at the same time. Similarly, with the significant interaction.

Most commonly, it will estimate the conditional expectation of the response of happiness given the risk factors. We also illustrate the percentage that the risk factors and the interactions contribute to the happiness score as we rank them and are shown below in Table 1.

**Table-1: Ranking of Individual Risk Factors and the Interactions With Respect to the Percentage of Contribution to the Response**

Rank	Risk Factors	Contributions(%)
1	LOG(GDP)	7.15
2	FREEDOM∩ PER_CORR	5.58
3	LOG(GDP) PER_CORR	5.00
4	FREEDOM	4.94
5	EXP(POS_AFFECT)	4.63
6	FREEDOM CONF_GOV	4.46

7	FREEDOM NEG_AFFECT	4.13
8	CONF_GOV SOC_SUPPORT	3.89
9	NEG_AFFECT SOC_SUPPORT	3.72
10	GENEROSITY	3.72
11	FREEDOM DEL_QUALITY	3.59
12	GENEROSITY DEL_QUALITY	3.45
13	EXP(SOC_SUPPORT)	3.30
14	GENEROSITY DEM_QUALITY	3.16
15	LOG(GDP) DEL_QUALITY	2.96
16	PER_CORR SOC_SUPPORT	2.87
17	LOG(GDP) LIFE_EXPECT	2.55
18	POS_AFFECT NEG_AFFECT	2.45
19	LOG(GDP) NEG_AFFECT	2.44
20	CONF_GOV POS_AFFECT	2.38

To assess the quality of the proposed analytical model [17] we use both the coefficient of determination,  $R^2$  and adjusted  $R^2$  ( $R^2_{adj}$ ) which are the key criteria to evaluate the model performance. The sum of square due to regression ( $SSR$ ), is the squared sum of the differences between the predicted response and the mean response. It captures the observed variability of the model. The sum of squared errors ( $SSE$ ), also termed as residual sum of squares, is the variation that remains unexplained. We always try to minimize this error in a model. The total sum of squares ( $SST$ ) =  $SSE + SSR$ .  $R^2$ , the coefficient of determination is defined as the proportion of the total response variation that is explained by the proposed model and it measures how well the regression process approximates the real data points. Thus,  $R^2$  is given by

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

However,  $R^2$  it does not consider the number of variables in the model. Also, there is the problem of the ever increasing  $R^2$ . To address these issues, we have the adjusted  $R^2$  which considers the number of parameters and is given by

$$R^2_{adj} = 1 - \left[ \frac{(1 - R^2)(n - 1)}{n - k - 1} \right],$$

Where  $n$  is the number of points in our data sample,  $k$  is the number of independent regressors, i.e., the number of risk factors in the model, excluding the constant. For our final statistical model, the R squared is **88.8%**, and R squared adjusted is 87.8%. Both R squared, and R squared adjusted is very high and very close to each other. That is, the developed statistical model explains 88.8% of the variation in the response variable, a very high-quality model. Similarly, the risk factor that we included in the model, along with the relevant interactions, estimates almost 89% of the total variation in the happiness score. In Figure 5, we rank the individual attributable variables and interactions (top 20 with respect to their contribution to the national

happiness score. That is, we listed those terms based on their percentage of contribution to the response. The ranking is important, given the fact that in a survey or experiment if the group of experimenters or surveyors know beforehand the most important variables which account for the response, they might collect information on those important variables only which will be economical and time-consuming as they might not be interested in the insignificant variables which contribute very minimum to the response or do not contribute at all.

**Verifying Model Assumptions**

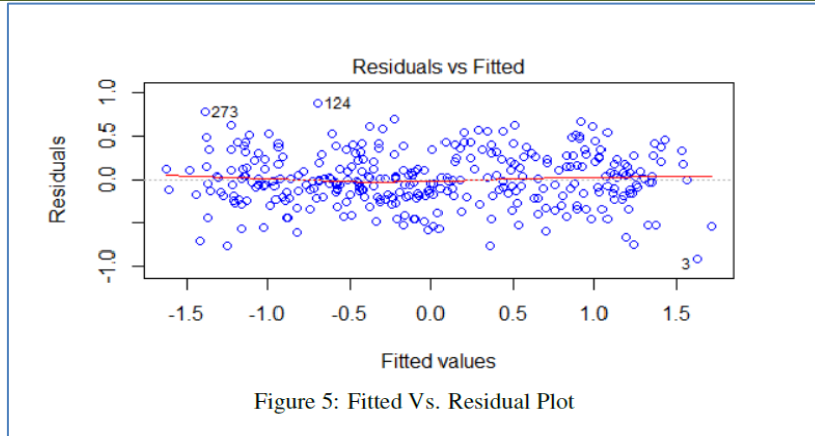
Once the statistical model has been developed, it is necessary to check the model assumptions (if any). In our case, we have proposed a multiple non-linear regression model, which is very useful and conveys to us accurately some important information on the subject matter. However, multiple linear regressions have some important assumptions which must be satisfied with the correctness of the proposed model. In this section, we will verify the important model assumptions.

**Residual Analysis**

When one performs multiple linear regression (or any other type of regression analysis), one obtains a linear function that best fits the data. The entire data points usually don't fall exactly on this regression plane but they are scattered around. A residual is a vertical distance between a data points and the regression plane. The residual (error) is defined as:

$$\hat{\epsilon} = residual = y - \hat{y},$$

Where  $y$  and  $\hat{y}$  are the observed and predicted response. The residual ( $\hat{\epsilon}$ ) is actually the estimated residual error from the linear fit. The sum of the residuals equals zero assuming that the regression line is actually the line of "best fit." In our case, the mean residual is  $-1.56 \times 10^{-18}$  implying that it is almost zero. Figure 5, below illustrates the behaviour of the residual estimator.



**Homoscedasticity of residuals**

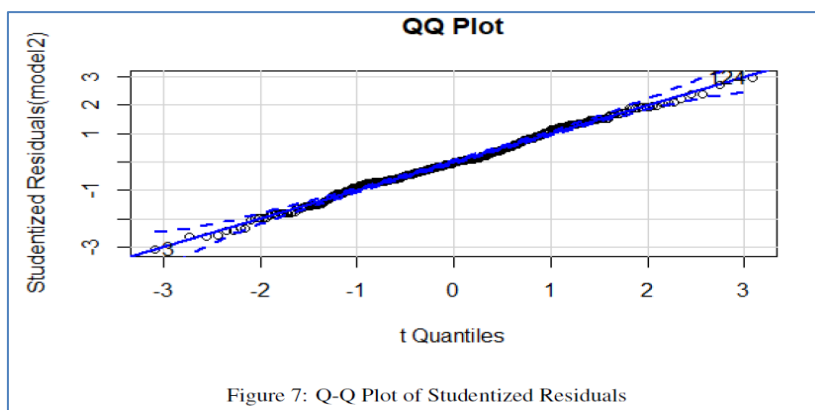
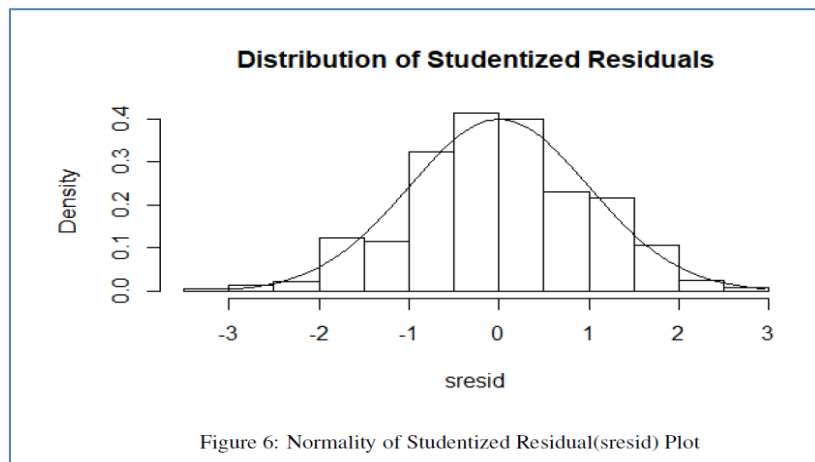
One of the main assumptions of the linear regression model is the homoscedasticity of the residuals or equal variance. That is,  $Var(\hat{\theta}) = \sigma^2$  which is constant. From Figure 5, we see that residuals vary as the fitted values increase. It seems that the pattern is more or less uniform, which is shown by the red line. There is no increasing or decreasing trend. Hence, the assumption of the constant variance of the residuals has been satisfied.

**Breusch-Pagan Test:** Breusch-Pagan (BPG) test is used to test for heteroskedasticity of the error

terms in a regression model. We obtained a p-value of .35173 by testing the null hypothesis of constant error variance against the alternative that the error variance changes with the level of the response (fitted values), or with a linear combination of predictors. Hence, we have significant reason to believe the error variance is constant.

**Normality of residual**

One important assumption of linear regression is normality of residual. From the following Figure 6 and Figure 7, we see that the studentized residual follows a normal pattern.





### No auto-correlation between the residuals

For linear regression models, it is expected to have zero auto-correlation between the error terms. That is,

$$\text{corr}(\hat{e}_i, \hat{e}_j) = \begin{cases} 0 & , \text{for } i \neq j \text{ and } i = j \\ 1 & \end{cases}$$

Where  $\hat{e}_i$  and  $\hat{e}_j$  are the  $i^{\text{th}}$  and  $j^{\text{th}}$  error terms in the model.

Figure 8 shows the autocorrelation of the residuals vs. lag plot. The X-axis corresponds to the lags of the residuals. The first line to the left shows the correlation of residuals with itself (Lag0); therefore, it will always be equal to 1. If the residuals were not auto-correlated, the correlation (Y-axis) from the immediate next line onwards would drop to a near-zero value below the dashed blue line (significance level). Hence, there is no auto-correlation between residuals in our model.

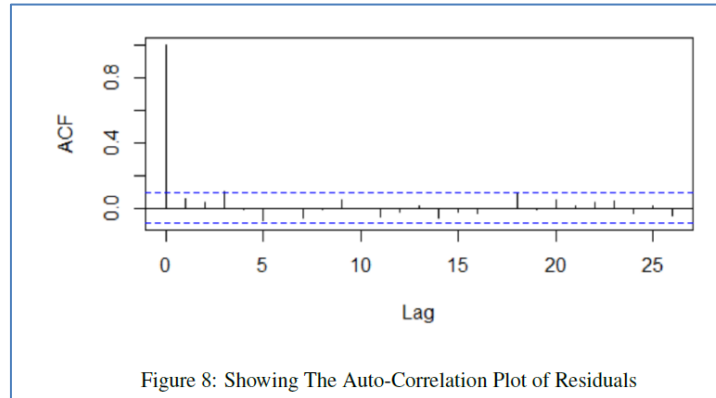


Figure 8: Showing The Auto-Correlation Plot of Residuals

**Run Test:** We can also verify the no auto-correlation case of the residuals by Run test (Wald, A. and Wolfowitz, J. (1940). Run test examines the randomness of a numeric sequence by studying the frequency of runs. We obtained a p-value of 0.9264, which implies that we fail to reject the null hypothesis that residuals are random. Hence, there is no pattern.

**Durbin-Watson Test:** The Durbin Watson Test is a measure of auto-correlation (also called serial correlation) in residuals from the regression analysis. Auto-correlation is the similarity of a timeseries over successive time intervals. It can underestimate the standard error and can cause one to believe that the predictors are significant when they are not. The Durbin-Watson test statistic is used to detect the presence of autocorrelation at lag 1 in the residuals (also termed as prediction errors) in regression analysis. The test statistic for this test is given by:

$$DW = \frac{\sum_{t=2}^T (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_t \hat{e}_t^2}$$

Where  $\hat{e}_t$  and  $\hat{e}_{t-1}$  are the residuals at time points  $t$  and  $t - 1$ , respectively.

A rule of thumb is that the test statistic values in the range of 1.5 to 2.5 are relatively normal. Values outside of this range could be a cause for concern. Field (2009) suggests that values under 1 or more than 3 are a definite cause for concern. The value we obtained for the test statistic is 1.89 with a p-value of .109, implying

that there is insufficient sample evidence to reject the null hypothesis that the true auto-correlation is zero.

### The regressors and the residuals are not correlated

We calculated the Pearson's product-moment correlation coefficient between each regressor and the residuals. As expected, every time we obtained an insignificant p-value implying that the true correlation is zero.

### Penalized Regression Models

Penalized regression methods have proven to be a high-yielding area of research in statistics and data sciences. The key idea is to add a 'penalty' to regression to encourage desirable behavior in the model. Often this is done to reduce variability in estimating the parameters. While developing the proposed statistical model for happiness, we used OLS, the ordinary least square technique to obtain an approximate estimate of the coefficients (weights) of the attributable variables. To address the multicollinearity problem (since in our data set, some variables are strongly correlated), the Regularization methods are used. Since these methods are based on adding the regularization parameters (lambda and alpha) to the regression coefficients of the individual risk factors, these the model generalizes the data and prevents over-fitting. To further illustrate our proposed model's quality, we will discuss three machine learning regularization methods and our proposed non-linear analytical model.

### Ridge Regression

For multiple linear regression, the ordinary least squares fitting procedure of the coefficient estimates (weights)  $\beta_1, \beta_2, \dots, \beta_p$  that minimizes the cost function RSS (Residual Sum of Squares), is given by,

$$RSS = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 .$$

Where  $\lambda \geq 0$  is a tuning parameter (sometimes called a penalty parameter that controls the strength of the penalty term in ridge regression) to be determined via cross validation.

$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{i=1}^p \beta_j^2 , \tag{5}$$

Ridge regression is very similar to least square regression, except that the ridge coefficients are estimated by minimizing a slightly different quantity. In particular, ridge regression coefficient estimates  $\widehat{\beta}_R$  are the values that minimize the following function:

**LASSO (Least Absolute Shrinkage and Selection Operator) Regression**

The LASSO regression method adds an absolute value of magnitude of a coefficient as penalty term to the loss function that can be expressed by:

$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{i=1}^p | \beta_j | , \tag{6}$$

Comparing (5) to (6), we see that the LASSO and Ridge regression have similar formulations.

LASSO uses an  $L_1$  penalty where the Ridge uses  $L_2$  penalty.

The only difference is that the  $\{\beta_j\}^2$  term in the ridge regression penalty in (6) has been replaced by in the LASSO penalty (6). In statistical parlance, the

**Elastic Net**

Elastic Net regression method which is the mix of Ridge and LASSO technique can be defined by:

$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \left[ (1 - \alpha) \sum_{i=1}^p \beta_j^2 + \alpha \sum_{i=1}^p | \beta_j | \right] , \tag{7}$$

Where  $\alpha$  is the mixing parameter between ridge ( $\alpha = 0$ ) and LASSO ( $\alpha = 1$ )

However, in the above equations (5, 6 and 7) the constructions of the three models will be the same structure as our proposed model in equation (1) with only the coefficient estimation will be different because of the randomness of choosing the training data set.

**Comparison among different Models**

We now proceed to compare the performance of the proposed model with the other three models using the following two matrices.

**Root Mean Squared Error (RMSE)**

After each repetition of the cross-validation, the model assessment metric RMSE is computed, which is given by:

$$RMSE = \left[ \frac{\sum_{i=1}^n (y_i - \widehat{y}_i)^2}{n} \right]^{\frac{1}{2}} ,$$

Where  $y_i$  and  $\widehat{y}_i$  are the observed and predicted responses.

**Mean Absolute Deviation (MAE)**

The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction which is given by:

$$MAE = \frac{\sum_{i=1}^n | y_i - \widehat{y}_i |}{n} ,$$

Where  $y_i$  and  $\widehat{y}_i$  are the observed and predicted responses.

While comparing the proposed model with the three regularization methods Ridge, LASSO, and Elastic Net, we have found that our proposed analytical model performs better in terms of validations matrices RMSE and MAE, as described above. Table 2 below, provides multiple comparisons among the different models in terms of training and testing accuracy.

**Table-2: Comparison Among Different Models in terms of RMSE and MAE**

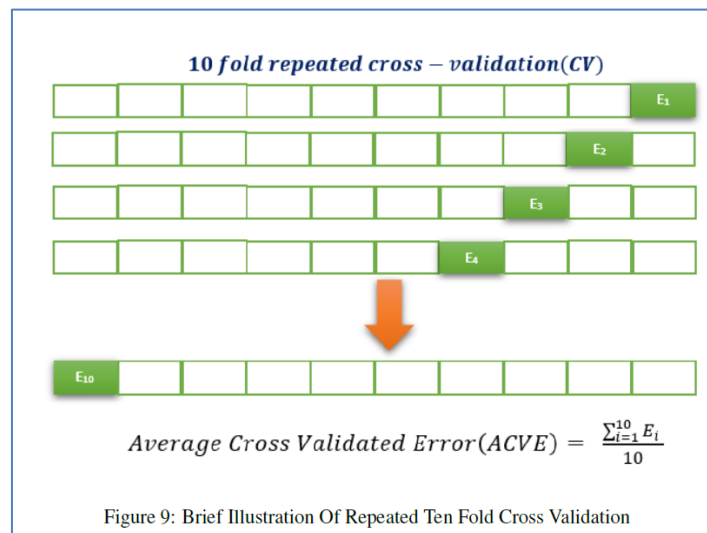
Table of compariosn				
Models	RMSE		MAE	
	Training	Testing	Training	Testing
<b>Proposed Model</b>	<b>.31</b>	<b>.43</b>	<b>.24</b>	<b>.31</b>
RIDGE	.38	.5	.3	.35
LASSO	.36	.52	.27	.37
Elastic Net	.36	.52	.29	.37

From the above Table 2, we see that our proposed nonlinear statistical model gives minimum testing error in terms of RMSE and MAE when compared with the penalized regression models. Thus, our analytical model outperforms the other three models for our happiness data.

**Validation and Prediction Accuracy of the Proposed Model**

We developed our analytical model on 80% training data and validated the model based on 20% testing data. In the testing data (validation data) the test error is the average error that occurs from using the analytical method to predict the response on a new observation. That is, a measurement that was not used in training the method. The test error gives an idea about the consistency of the analytical model. Moreover, we performed repeated ten-fold repeated cross-validation (10 times) for our validation testing.

The primary objective is that we will use 10-fold cross-validation, and then we repeated cross-validation ten times, where each of the repetition folds are split differently. In 10-fold cross-validation, the training set is divided into ten equal subsets. One of the subsets is taken as the testing set in turn, and (10-1) = 9 subsets are taken as a training set in the proposed model. The error mean square error  $E_1$  is computed for the held out set. This procedure is repeated 10 times; each time, a different group of observations is treated as a validation set. This process results in 10 estimates of the test error,  $E_i; i = 1, 2, \dots, 10$ . The average error of each set, throughout the cross-validation process is termed as a cross-validated error. The following Figure 9, illustrates briefly the idea of 10 fold repeated cross-validation, where  $E_i; i = 1, 2, \dots, 10$  is the mean square error (MSE) in each iteration and ACVE is the average cross-validated error.



Now we employ the following two methods to illustrate the prediction performance of the proposed model.

**Min-Max Accuracy**

Min-Max-Accuracy is the average of the ratio of minimum value between the actual observation and predicted observation and maximum between actual

observation and predicted observation. Mathematically, it can be expressed as follows:

$$Min - Max - Accuracy = Mean \left[ \frac{\min(y_i, \hat{y}_i)}{\max(y_i, \hat{y}_i)} \right]$$

Where  $y_i$  and  $\hat{y}_i$  are the observed and predicted responses.

It gives an idea about how far the model’s prediction is off on an average. For a perfect model, this measure is 1. This can be taken as the accuracy of the proposed model. For our developed model, the Min-Max accuracy is 96.2%, which is quite impressive.

**Correlation Accuracy**

A simple correlation between the original observations and predicted observations can be used as a form of accuracy measure. A greater correlation accuracy implies that the original and predicted observations have analogous directional movement, i.e., when the original observations increase, the predicted observations also increase and vice-versa. We obtained a correlation accuracy of 90.5% in the test data, which implies that our statistical model is of high quality and should be useful for applied predictive analysis for real data.

Table 3 below, provides the two measures of prediction accuracy for our proposed model.

**Table-3: Prediction Accuracy for Proposed Model**

Min-Max-Accuracy	Correlation Accuracy
96.2%	90.5%

Thus, the above two methods attest to the high quality of our proposed model.

**RESULTS AND DISCUSSION**

After obtaining the significant risk factors along with their significant interactions, we rank them with respect to the percent of contribution to the happiness scores for the developing countries as shown by Table 1. The risk variable that has the largest contribution to the happiness score is the variable LOG (GDP) which contributes 7.15% of the total variation to the happiness score. The next largest contribution is the

combined effect of freedom and perception of corruption with 5.58% contribution. Numbers 3, 4, and 5 are the combined interaction effect of LOG (GDP), FREEDOM, and exp(POS\_AFFECT) with a contribution of 5%, 4.94%, and 4.63%, respectively. Hence, summing these risk factors up, we identify that they explain almost 89% of the total variability in the national average happiness score for all developing countries. We can address the usefulness and importance of the proposed model in the subject area in five important categories.

These categories are given below.

1. We have identified and tested the individual attributable variables (risk factors) responsible for the change in happiness score across all the developed countries.
2. We have identified the significant interactions that influence the happiness score, in our model.
3. We have ranked the individual risk factors and interactions as a percentage of contribution for the response of the national average of happiness score (Ladder) or subjective well-being (SWB).
4. We can obtain excellent predictions of happiness of individuals given the values of the attributable variables from our analytical model with a high degree of accuracy.
5. Any particular country might use our non-linear statistical model to work on specific risk factors to increase their happiness score. As for example, one can work on the variable SOC\_SUPPORT ( $X_2$ ) if the value for a particular year is not satisfactory and work on other important aspects to increase the value so that the happiness score can be increased.

We have also ranked all the developed countries based on the predicted happiness score of the most recent observations (data) available for the year 2019. The following Table 4, illustrates the ranking of the countries.

**Table-4: Ranking of Developed Countries based on Predicted Happiness Score**

Rank	Country	Score	Rank	Country	Score
1	Finland	7.67	28	Belarus	6.51
2	Denmark	7.55	29	Belgium	6.51
3	Sweden	7.54	30	Czechia	6.46
4	Iceland	7.38	31	Norway	6.43
5	USA	7.35	32	Israel	6.38
6	Canada	7.29	33	Lithuania	6.35
7	Ireland	7.17	34	Chile	6.34
8	Switzerland	7.16	35	Spain	6.31
9	UK	7.09	36	Slovakia	6.24
10	Germany	7.03	37	Japan	6.24
11	Malta	6.98	38	Hungary	6.14
12	Luxembourg	6.96	39	Poland	6.12
13	Oman	6.96	40	New-Zealand	6.11
14	Estonia	6.92	41	Cyprus	6.09
15	Singapore	6.89	42	Italy	6.06

16	Qatar	6.82	43	Kazakhstan	6.05
17	France	6.76	44	Russia	5.99
18	Uruguay	6.73	45	South Korea	5.95
19	Slovenia	6.62	46	Kuwait	5.70
20	Malaysia	6.61	47	Turkey	5.60
21	UAE	6.56	48	Croatia	5.55
22	Saudi Arabia	6.54	49	Portugal	5.45
23	Netherlands	6.53	50	Montenegro	5.38
24	Argentina	6.51	51	Latvia	5.35
25	Australia	6.51	52	Bulgaria	5.34
26	Austria	6.51	53	Greece	5.20
27	Bahrain	6.51	54	Romania	5.04

It is interesting to note that Finland and Denmark possess the top happiness scores while the United States is fifth. Also, studies [19], has shown a significant influence of democracy on an individuals' subjective well-being (happiness). Finland and Denmark falling into the category of the top democratic countries of the world, also validates the fact.

## CONCLUSION

We have developed a real data-driven analytical model that very accurately identifies the following very useful findings concerning the happiness of the society of developed countries in the world:

- Identifies the significant attributable variables (risk factors) that drive the degree of happiness.
- Identifies the significant interactions of the risk factors that contribute to the degree of happiness.
- The developed analytical model that predicts the degree of happiness very accurately for a given response to a set of questions.
- The developed model can be used strategically to increase the degree of happiness by working with the identified risk factors.
- Furthermore, one can perform surface response analysis to identify the target values of the risk factors so as to be, say, 95 percent sure that we will maximize the degree of happiness based on the identified values.

The developed analytical model has been evaluated by several statistical methods that include the  $R^2$  and  $R^2_{adjusted}$  that attest to its high quality. The risk factor GDP is the highest contributor to the happiness score contributing 7.15%, while DEL\_QUALITY contributes the least with 1.31% to the response. The findings of our study suggest that economists and other social scientists might need to pay more attention to emotional well-being as a causal force. Also, since individual happiness in an organization has a positive correlation with productivity, our proposed statistical model can be used for firms' promotion policies, and they may be useful for managers and human resources professionals. Human resource managers can use our model to predict the individual happiness score by using the questionnaire (attached in appendix 1). It will help the company to identify those individuals who need to be rewarded and those who need to improve their

happiness score. Identifying those individuals is essential for the company as happiness is correlated with the increase in productivity. Our proposed statistical model is also highly useful for decision making and strategic planning on controlling the factors responsible for causing people to be unhappy and depressed. Finally, since happiness is the most crucial aspect of human life that we seek for, controlling the most critical risk factors that significantly contribute to the happiness are essential to control the crime rate of a country, as there is a negative correlation between the individual country's happiness score (Ladder) and crime rate.

## Availability of data and material

The data can be obtained online from the source (<https://worldhappiness.report/ed/2019/>).

## REFERENCE

1. Oswald AJ, Proto E, Sgroi D. Happiness and productivity. *Journal of Labor Economics*. 2015 Oct 1;33(4):789-822.
2. Perneger TV, Hudelson PM, Bovier PA. Health and happiness in young Swiss adults. *Quality of Life Research*. 2004 Feb;13(1):171-8.
3. DeNeve KM, Cooper H. The happy personality: a meta-analysis of 137 personality traits and subjective well-being. *Psychological bulletin*. 1998 Sep;124(2):197.
4. Pettijohn TF II, Pettijohn TF. Perceived happiness of college students measured by Maslow's hierarchy of needs. *Psychol Rep*. 1996; 79: 759-762.
5. Frey BS, Stutzer A. Happiness prospers in democracy. *J Happiness Studies*. 2000; 1: 79-102.
6. Myers DG, Diener E. The pursuit of happiness. *Sci Am*. 1996; 274: 54-56.
7. Cammock T, Joseph S, Lewis CA. Personality correlates of scores on the Depression-Happiness Scale. *Psychol Rep*. 1994; 75: 1649-1650.
8. George MS, Ketter TA, Parekh PI, Horwitz B, Herscovitch P, Post RM. Brain activity during transient sadness and happiness in healthy women. *Am J Psychiatry*. 1995; 152: 341-351.
9. Koivumaa-Honkanen H, Honkanen R, Viinama ki H, Heikkila K, Kaprio J, Koskenvuo M.



- Selfreported life satisfaction and 20-year mortality in healthy Finnish adults. *Am J Epidemiol*. 2000;15: 983–991.
10. Berwick DM, Murphy JM, Goldman PA, Ware JE, Barsky AJ, Weinstein MC. Performance of a five-item mental health screening test. *Med Care*. 1991; 29: 169–176.
  11. McDowell I, Praught E. On the measurement of happiness. An examination of the Bradburn scale in the Canada Health Survey. *Am J Epidemiol*. 1982; 116: 949–958.
  12. Bradburn N. *The Structure of Psychological Well-Being*. Chicago: Adline; 1995.
  13. Walsh J, Joseph S, Lewis CA. Internal reliability and convergent validity of the Depression–Happiness Scale with the General Health Questionnaire in an employed adult sample. *Psychol Rep*. 1995; 76: 137–138.
  14. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. 1970 Feb 1;12(1):55-67.
  15. Tibshirani R. Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Methodological)*. 1996; 58: 267–288.
  16. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*. 2005 Apr 1;67(2):301-20.
  17. Sheha MA, Tsokos CP. Statistical Modeling of Emission Factors of Fossil Fuels Contributing to Atmospheric Carbon Dioxide in Africa. *Atmospheric and Climate Sciences*. 2019 May 20;9(3):438-55.
  18. Polansky AM, Chou YM, Mason RL. An algorithm for fitting Johnson transformations to non-normal data. *Journal of quality technology*. 1999 Jul 1;31(3):345-50.
  19. Dorn D, Fischer JA, Kirchgässner G, Sousa-Poza A. Is it culture or democracy? The impact of democracy and culture on happiness. *Social Indicators Research*. 2007 Jul 1;82(3):505-26.
  20. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC medical research methodology*. 2014 Dec;14(1):1-3.
  21. Rubin DB. Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics*. 1986 Jan 1;4(1):87-94.
  22. Helliwell J, Layard R and Sachs J. *World Happiness Report 2019*, New York: Sustainable Development Solutions Network. 2019; (<https://worldhappiness.report/ed/2019/#read>)
  23. Thomas M. What Do the Worldwide Governance Indicators Measure?. *Eur J Dev Res*. 2010; 22: 31–54. <https://doi.org/10.1057/ejdr.2009.32>
  24. Wald A, Wolfowitz J. On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*. 1940 Jun 1;11(2):147-62.
  25. White, Kenneth J. “The Durbin-Watson Test for Autocorrelation in Nonlinear Models.” *The Review of Economics and Statistics*. 1992; 74(2): 370–373. JSTOR, [www.jstor.org/stable/2109675](http://www.jstor.org/stable/2109675). Accessed 9 Dec. 2020.
  26. Halinski RS and Feldt LS. The selection of variables in multiple regression analysis. *Journal of Educational Measurement*. 1970; 7: 151-157. <https://doi.org/10.1111/j.1745-3984.1970.tb00709.x>

**Appendix I**

**SURVEY QUESTIONNAIRE**

Based on who is requesting the information for an individual that is associated with any Government, Company, Organization, Educational Institutions, etc.

**GDP(X1):** Per-capita gross domestic product of the country the individual resides (given information)

**Social Support(X2):** If you were in trouble, do you have relatives or close friends, you can count on to help you whenever you need them? A) YES  B) NO

**Life Expectancy(X3):** From the attached graph, identify your life expectancy.  Years.

**Freedom(X4):** Are you satisfied with your freedom to choose what you do with your life?  
A) YES  B) NO

**Generosity(X5):** Have you donated money to a charity in the past month? A) YES, B) No. If the answer is YES, then how much?

**Corruption Perception(X6):** Is corruption widespread throughout your government, your company, or your organization? A) YES  B) NO

**Positive Affect(X7):** Happiness, laughter, and enjoyment.

7.1. On a scale of 1 to 10, how **happy** were you for the last five days?

7.2. On a scale of 1 to 10, how much did you **laugh** for the last five days?

7.3. On a scale of 1 to 10, how much did you **enjoy** for the last five days?

**Negative Affect(X8):** Worry, Sadness, and anger, respectively.

8.1. On a scale of 1 to 10, how **worried** were you for the last five days?

8.2. On a scale of 1 to 10, how **sad** were you for the previous five days?

8.3. On a scale of 1 to 10, how **angry** were you for the last five days?

**Confidence in Government(X9):** In your government, company, or organization, etc. how much trust and confidence do you have when it comes to handling [International problems/Domestic problems]?

A) A great deal  B) A fair amount   
C) not very much  D) none at all

**Democratic Quality (X10):**

10.1. On a scale of 1 to 10, how likely do you think that the country's citizens can participate in selecting their government, enjoy Freedom of expression, Freedom of association, and unprejudiced media coverage?

Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10.2. On a scale of 1 to 10, how likely you think people suffer consequences of political instability and politically motivated violence, including terrorism?

Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Delivery Quality(X11):**

11.1. On a scale of 1 to 10, how likely do you think that your company/organization/government has maintained the quality of public services, the quality of the civil service, the quality of policy formulation and implementation, and the credibility of such policies?

Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11.2. On a scale of 1 to 10, how likely do you think that your company/organization/government can formulate and implement sound policies and regulations that permit and promote private sector development?

Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11.3. On a scale of 1 to 10, how likely do you think that your company/organization/government agents and law enforcement agencies have confidence in the government and abide by society's rules?

Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11.4. On a scale of 1 to 10, to what extent you think that public power is exercised for private gain, including both petty and grand forms of corruption by the elites for their individual interests?

Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10.2. On a scale of 1 to 10, how likely you think people suffer consequences of political instability and politically motivated violence, including terrorism?

Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Delivery Quality(X11):**

11.1. On a scale of 1 to 10, how likely do you think that your company/organization/government has maintained the quality of public services, the quality of the civil service, the quality of policy formulation and implementation, and the credibility of such policies?

Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11.2. On a scale of 1 to 10, how likely do you think that your company/organization/government can formulate and implement sound policies and regulations that permit and promote private sector development?

Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11.3. On a scale of 1 to 10, how likely do you think that your company/organization/government agents and law enforcement agencies have confidence in the government and abide by society's rules?

Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11.4. On a scale of 1 to 10, to what extent you think that public power is exercised for private gain, including both petty and grand forms of corruption by the elites for their individual interests?

Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>