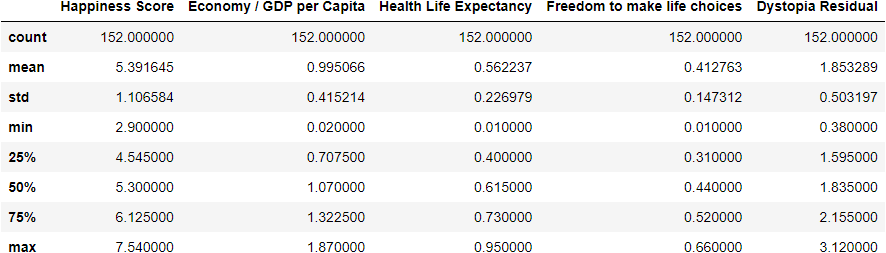
* 1. **Data set**

The data used in study includes four independents’ variables with Happiness Score collected from 152 countries in 2017. The data set is collected by Gallup World Poll by calculating scores based on the sampled people’s answers to the key life evaluation. The happiness scale of 0 to 10 is equivalent to the increase in their rate.

The variables used in multiple linear regression models are Happiness Score, Economy / GDP per Capita, Health Life Expectancy, Freedom to make life choices and Dystopia Residual. In which, Dystopia is an imaginary country whose values equivalent to the world’s averages for each of the four variables. The Dystopia Residual factor is the Dystopia Happiness Score (1.85) + the Residual value or the unexplained value for each nation as mentioned in the front answer.

* 1. **Preliminary analysis and descriptive statistics**

Table 1 presents our calculations on the mean, maximum, minimum, standard deviation, variance values of each variables.





Looking at the detail, standard deviation values of four dependent variables are small which ranges from 0.147312 to 0.503197, which indicates the accuracy of the results presented in the index. Regarding the Mean index, we can see that mean of Dystopia Residual is 1.853289 which is highest compared to the mean of other variables. Therefore, the temporary conclusion is that Dystopia Residual is the most important factor to evaluate happiness scores which present one-third in total Happiness scores, whereas there was just only 7.64% of score presented by Freedom.

* 1. **Graphical presentation of data**

Charts below present the relationship between happiness score and each independent variable. Each dot represents variable value of each country.

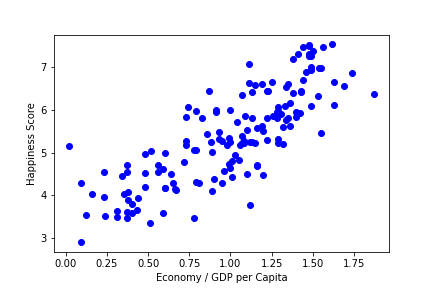


Figure 1: Scatter plot between GDP per capita and Happiness Score



Figure 2: Scatter plot between Health Life Expectancy and Happiness Score

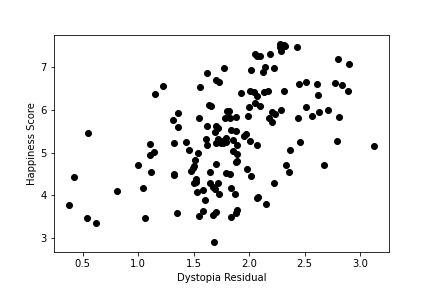


Figure 3: Scatter plot between Dystopia Residual and Happiness Score

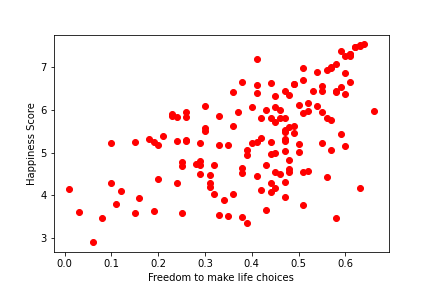


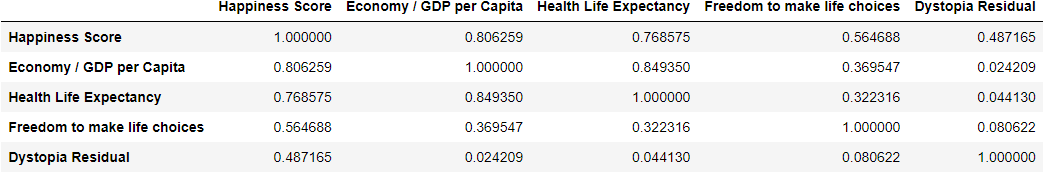
Figure 4: Scatter plot between Freedom to make life choices and Happiness Score

In details, in Figure 1, the scatter plot shows the relationship between GDP per capita and Happiness Score distribution. Each dot represents GDP per capita of each country. From the plot, we can see a positive correlation between GDP per capita and Happiness Score; when GDP increase, Happiness Score also increase.

Similarly, figure 2, figure 3 and figure 4 also show positive correlation between Happiness score and each of independent variables. In other word, the increase in value of each factor will lead to the increase of Happiness score.

* 1. **Adequate test’s results**

Table 2 presents the correlation between all variables.



The results of correlation reported in Table 2, present evidence of correlation analysis between all determinants variables, meaning that the happiness score variable is correlatated with other independant variables. From the Pearson’s correlation outcomes and the line of scatterplot charts showed above, it can be interpreted that there are positive correlations among studied variables. In details, Economy/GDP per Capita and Health/ Life Expectancy have strong correlation with Happiness Score, with the Pearson’s correlation outcomes of 0.806259 and 0.768575, respectively. Besides, Freedom to make life choices and Dystopia Residual have medium tight correlation with Happiness Score, with the Pearson’s correlation outcomes of 0.564688 and 0.487165 respectively. Among four other independent variables, the correlation between Happiness Score and Economy/GDP per Capita is considered as the strongest.

* 1. **Model Summary and Model:**

Table 2 below presents the regression results.

**OLS Regression Results**

==================================================================

|  |  |  |
| --- | --- | --- |
| Dep. Variable: y |  | R-squared: 0.950 |
| Model: OLS |  | Adj. R-squared: 0.949 |
| Method: Least Squares |  | F-statistic: 703.1 |
| Date: Thu, 07 Apr 2022 |  | Prob (F-statistic): 1.04e-94 |
| Time: 16:18:22 |  | Log-Likelihood: -2.4000 |
| No. Observations: 152 |  | AIC: 14.80 |
| Df Residuals: 147 |  | BIC: 29.92 |
| Df Model: 4 |  |  |
| Covariance Type: nonrobust |  |  |

==================================================================

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | coef | std err | t | P>|t| | [0.025 | 0.975] |
| const | 0.78160 | 0.099 | 7.926 | 0.000 | 0.587 | 0.977 |
| x1 | 1.2771 | 0.095 | 13.494 | 0.000 | 1.090 | 1.464 |
| x2 | 1.2437 | 0.170 | 7.317 | 0.000 | 0.908 | 1.580 |
| x3 | 2.0259 | 0.149 | 13.591 | 0.000 | 1.731 | 2.320 |
| x4 | 0.9732 | 0.041 | 23.973 | 0.000 | 0.893 | 1.053 |

==================================================================

|  |  |  |
| --- | --- | --- |
| Omnibus: 2.019 |  | Durbin-Watson: 1.681 |
| Prob(Omnibus): 0.364 |  | Jarque-Bera (JB): 1.570 |
| Skew: -0.212 |  | Prob(JB): 0.456 |
| Kurtosis: 3.262 |  | Cond. No. 23.1 |

==================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Table 2: Model Results

* ***The Coefficient of Multiple Determination (R2)***

The coefficient of multiple determination shows that the percentage of variation in the dependent variable (Y) that is interpreted by the variance in the set of independent variables (X) in a multiple regression model (Berenson et al. 2015). This large R2 in our model demonstrates that 95% of the variability in the happiness score is explained by the variability in a variety of factors, including Freedom, Health life expectancy, GDP per Capita, and Dystopia Residual. Alternatively, only 5% of the variation is due to other factors.

* ***Adjusted R-Squared***

When analysing a multiple regression model, the calculation process will inflate the R-squared. Therefore, it is necessary to utilise the adjusted R2 to consider both the sample size and the set of independent variables in the model. According to the model summary table, 94,9% of the fluctuation in happiness score is interpreted by the multiple regression model - adjusted for the sample size and the number of independent variables.

* ***Sample model***

From the results in Table 2, the multiple regression model is:

**Ŷi­ = 0.7816 + 1.2771X1 + 1.2437X2 + 2.0259X3 + 0.9732X4**

The regression coefficients are interpreted as follows:

* The regression constant β0is 0.7816. This means that if all four criteria are scored with 0, the estimated happiness score will be 0.7816.
* The regression coefficient of Economyis 1.2771. It means that holding constant the impact of 3 factors including healthy life expectancy, freedom and dystopia residual, for each increase of 1 point in the economy score, the happiness score will be predicted to increase by 1.2771 points. Therefore, the economic factor has a positive effect on the happiness score.
* The regression coefficient of Healthy lifeis 1.2437. This means that holding constant the influence of 3 criteria economy, freedom and dystopia residual, the predicted happiness score raises by 1.2437 points for each increase of 1 point in the healthy life expectancy score.
* The regression coefficients of Freedom and Dystopia Residual are 2.0259 and 0.9732, respectively. It is evident that both Freedom and Dystopia Residual have a statistically significant influence on happiness scores. In detail, Freedom has the most positive effect on happiness score, compared to the remaining three factors. The figure for Freedom also indicates that for every one-unit increase in Freedom, the dependent variable (happiness score) increases by 2.0259 units. Also, similarly, for Dystopia Residual, for each unit increase in Dystopia Residual, the dependent variable increases by 0.9732 units.