

# Cross-country analysis on connection between Financial Lifestyle and Happiness

Anastasiia Dziuba  
Computer Science  
(Master Program)  
Faculty of Mathematics, Natural Sciences  
and Information Technology  
(University of Primorska)  
Koper, Slovenia  
[89222003@student.upr.si](mailto:89222003@student.upr.si)

Uroš Sergaš  
Department of Information Sciences and  
Technologies  
Faculty of Mathematics, Natural Sciences  
and Information Technology  
(University of Primorska)  
Koper, Slovenia  
[uros.sergas@upr.si](mailto:uros.sergas@upr.si)

**Abstract:** The aim of this research is to investigate if happiness depends on the lifestyle inhabitants can allow themselves from financial point of view. Furthermore, this paper examines in which countries and regions it is possible to have middle class expenses with average salary. In addition, correlation between other features and happiness is observed. For specific correlations, linear models and clustering techniques are presented to identify associations. Obtained results and patterns behind them are further discussed which provides clearer understanding of research topic.

**Keywords:** happiness, cost of living, financial condition, monthly expenses

## 1. Introduction

Nowadays term “happiness” is getting more widely used. In this century Science of well-being is one of the most “mysterious” ones. There are different assumptions regarding the reason behind being happy. For instance, Brooks provides different scientific claims regarding the “ingredients of happiness”. Firstly, happiness is described as subjective well-being, which is comprised of genes, circumstances and habits. Habits, according to the author, are comprised from faith, family, friends and work. Brooks believes that work in this equation, means that the individual can earn success and serve others. However, he claims, that: “Happiness cannot buy

satisfaction.” Since people can never have enough of money, their wants are rising with paycheck and picture of perfect life is adjusted very quickly. (Brooks, 2022).

Kahneman argued in his study from 2010, that income influences life satisfaction, but not the happiness (Kahneman & Deaton, 2010). However, in 2023, Killingsworth, Kahneman and Mellers in their article “Income and emotional well-being: conflict resolved”, concluded that higher income makes people happier even in the high ranges of income. However, the authors also mentioned that the bottom of happiness distribution rises much faster than the top in that range of incomes. (Killingsworth et al., 2023)

Nevertheless, the role of money is considerably rising in society. We can notice that now there is a trend of “buying happiness”. This century can be described with consumerism and of course it makes an imprint on the people understanding of happiness. Social media influences it as well, we try to compare our lifestyle with others, it is starting from the small age. Before people did not have that much of entertainment and could value small things. Now it is impossible to limit our wants.

The aim of this research is to observe how financial condition of inhabitants varies from country to country taking into account average expenses for developed countries. Nowadays we can find various indexes of evaluating the life on internet. These happiness scores are based on different features. In this research we will consider the index provided by World Happiness Report.

## 2. Methodology and Gathering data

In order to study inhabitants' happiness score and their financial condition, exploratory data analysis was applied. For this purpose, the quantitative data was used. The data for this research was obtained from several sources since I could not find recent data on needed features on one source.

Therefore, I used Gapminder

(<https://www.gapminder.org>) , Kaggle

(<https://www.kaggle.com>) and WHO

(<https://www.who.int>) . Kaggle is a site meant for machine learning practitioners and data scientists, Gapminder is a site displaying time series of development statistics for all countries, WHO stands for World Health Organization, specialized agency of United Nations responsible for international public health.

For this research I used the following regions classification, which includes Australia and Oceania, North America, South America, Central America and Caribbean, Europe, East and Southeast Asia, Central Asia, South Asia, Middle East and Africa.

To omit bias while using the data from different resources, only the common values were used for observation. More details regarding data cleaning will be provided further for the specific datasets. Every dataset I used was available with . csv extension and was published to public domain and is free to use.

### A. Cost of living dataset

Dataset was provided on Kaggle by Miguel Piedade, the member and contributor of platform. Miguel Piedade gathered the data from Numbeo website (<https://www.numbeo.com/cost-of-living/>) by web scraping. Numbeo is world's cost of living database, where the data is updated on quarterly basis. The dataset published on Kaggle was based on prices in 3<sup>rd</sup> quarter of 2022. The dataset comprised of 4874 cities from different countries, due to data quality, this list was reduced to 760 cities. However, I used only capital cities for observation, which ended up to 133 different countries. The data provided on Kaggle comprised of 54 price metrics, in my research I used following metrics as average required expenses for living: internet, fitness club (monthly fee), monthly pass (local transport), basic (electricity, heating, cooling, water, garbage), 550 min of prepaid mobile tariff local, 4 meals in inexpensive restaurant, 8 bottles of water, 24 cups of coffee, 2 pieces of clothes. In addition, the expenses for

food were based on the minimum basket list for consuming 2400 calories daily and using western diet type, which can be found here (<https://www.numbeo.com/food-prices/>) . Firstly, I included "rent expenses" as well, however, it led to biased results. Since a lot of people live not alone, it is difficult to evaluate the rent price for individual. We should also consider that some part of population lives in their own properties, therefore, they do not have monthly rent expenses.

### B. Happiness score dataset

Dataset was published as an appendix to figures for World Happiness Report 2023

(<https://worldhappiness.report/>) , which is a publication of the Sustainable Development Solutions Network. The data for the report is powered by Gallup World Poll, global analytics and advice company (<https://www.gallup.com/>) .

The company collects data using telephone surveys in majority of cases, however, in some parts of the world (Latin America, Asia, Africa, etc.) face-to-face interviewing in randomly selected households. The happiness rankings are based on the participants' own assessment of their lives, in particular a single-item Cantril ladder life-evaluation question (from 1 to 10, where 10 is being the highest satisfaction with life). Nevertheless, the data was published in report for 2023, the survey was conducted in 2022.

### C. Other datasets

In order to see if other features impacting the life evaluation of people worldwide, I used following datasets: Life Expectancy, Human Development Index, Coverage Index for essential health services (UHC index), number of suicides, which was further adjusted to suicide rate (number of suicides per capita), child mortality rate, GDP per capita, Net migration. Most of the data was released in 2021, however, the datasets for child mortality rate and number of suicides were lastly updated in 2019. I obtained the datasets for Life expectancy, UHC index and Child mortality rate from World Health Organization. Net migration rate was gathered from Central Intelligence Agency. GDP per capita was included in the dataset published as annex for World Happiness Report. HDI (Human Development Index) was published in Human Development Reports.

The final dataset with features abbreviations and description that was used for analysis is provided below in Table 1.

Table 1

Column name:	Measurement:	Variable type:	Description :
country	categorical	nominal	Name of the country
region	categorical	nominal	Region to which country belongs
savings	numerical	ratio	Average salary after deducting minimal expenses and taxes (in US dollars)
happiness	categorical	ordinary	Metric measured in 2022 by asking participants about their life satisfaction (the scale from 1 to 10)
life	numerical	ratio	Life Expectancy, estimate of average age of participants
hdi	categorical	ordinary	Human Development Index, summary measure of average achievement in key areas of human development
hcov	categorical	ordinary	Coverage index for essential health services
suicide	categorical	ordinary	Suicide rate, number of suicides per

			100000 people
child	categorical	ordinary	Child mortality rate, probability that a child will die in the period between birth and 5 years
gdp	numerical	ratio	GDP per capita (in US Dollars)
mig	numerical	ratio	Net migration, difference between immigration and emigration in particular country

### 3. Statistical Analysis and its Results

#### i. Heatmaps

Firstly, let us observe the situation regarding cost of living in the world. Using heatmaps, it is clearly seen in which countries inhabitants are left with negative balance after average spendings (from red to light red) and where people have opportunity to save some money (shades of blue). However, it can be complicated to look at the specific country, therefore, plots of all world parts are provided below.

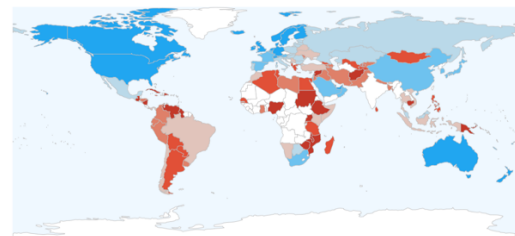
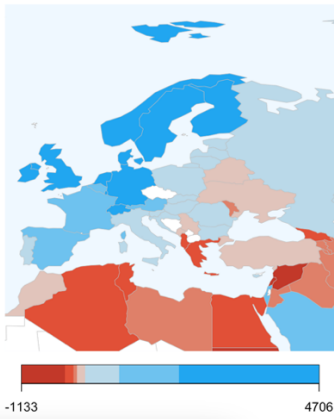
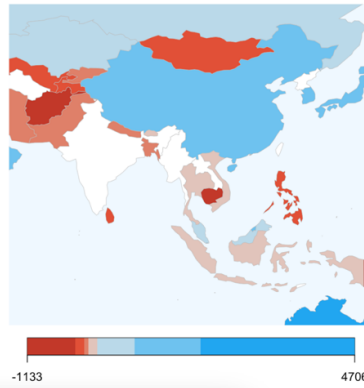


Figure 1

Account balance at the end of the month in Europe



Account balance at the end of the month in Asia



Account balance at the end of the month in Latin America

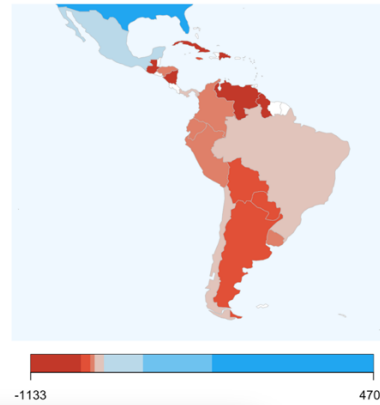
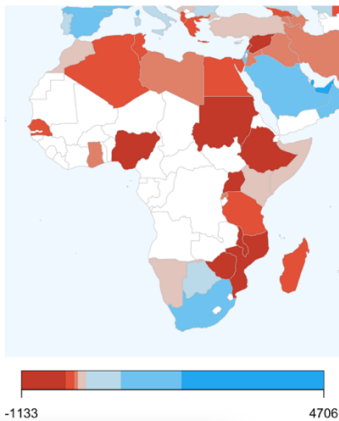
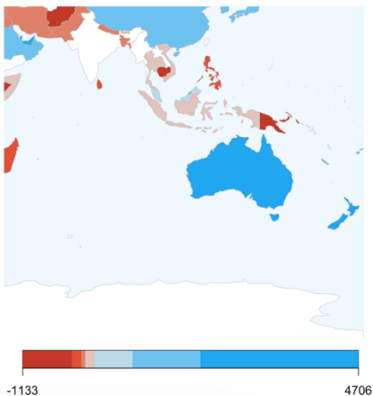


Figure 2

Account balance at the end of the month in Africa



Account balance at the end of the month in Oceania



## ii. Regions' outlook

Figure 3 shows us distributional characteristics of groups of savings based on the region classification. We can notice different patterns in the world regions relying on the shapes and positions of boxplots. Boxplot for Africa is comparatively short, which suggests that country values there, are low dispersed and normally distributed. On average people there cannot afford all expenses. Australia and Oceania show normal distribution and low dispersion as well, however, the average there is surplus of 3000\$. Boxplot for East and Southeast Asia suggests positively skewed distribution of values, therefore there is higher dispersion in upper quartile. Nevertheless, on average people in this region are left with almost nothing at the end of the month (median  $\approx 0$ ). Let us look at the European region, where the values are also highly dispersed and right skewed distribution can be observed. We can suggest that the reason for this lies in different parts of European region, since in the used countries classification there was not any division between countries there. There is less dispersion in 1<sup>st</sup> and 2<sup>nd</sup> quartiles of boxplot, it is interesting that the maximum values go beyond 4000\$ (the 4<sup>th</sup> quartile shows high dispersion of values), however, the median is around 750\$.

Eastern countries share similar distribution pattern as European ones, although, the account balance varies there from negative to 1500\$. We can notice high dispersion in 3<sup>rd</sup> quartile and significantly low one in 2<sup>nd</sup> quartile. Despite considerably high maximum, on average people in Middle Eastern countries do not have money for savings at the end of the month. Distribution of Northern American region tends to normal, therefore, people there are left with 2000\$ on their bank account. South Asian and South American countries share similar pattern; however, we can notice negative outlier in South Asia. People in both regions do not have enough money to afford average level of life. Unfortunately, outlook of the Central Asian region is quite biased, only 2 countries are present in this group due to the low data quality.

Overall, we can observe that the financial situation is quite different in world regions, which demonstrates inequality between developing and developed countries regarding the level of industrialization.

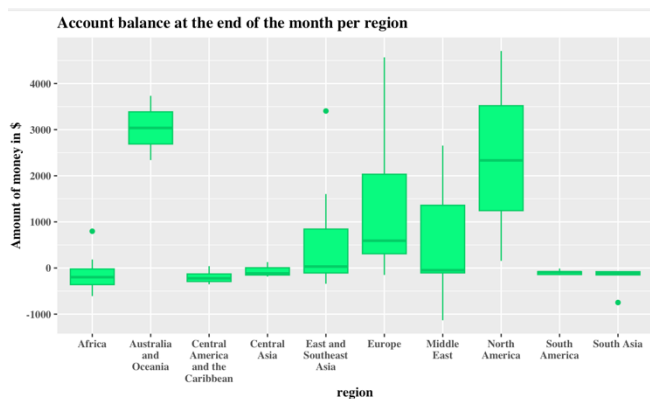


Figure 3

### iii. Correlation

Correlation plot gives us opportunity to quantify the direction and strength of the relationship between included in our research variables. Figure 4 gives us the outlook of correlation in our dataset. We can notice that there is high positive correlation between happiness ranking and health coverage, human development index, life expectancy, gdp and financial condition of people. In addition, there high negative correlation between happiness score and child mortality. However, from the plot we assume that there is low negative correlation between suicide rate and other presented variables. On the other hand, relationships between financial condition and other variables shows similar picture as happiness score. The difference is noticeable between life expectancy and financial

condition, where the correlation is moderate positive. Finally, we observe low negative correlation between financial condition and child mortality.

Overall, we can assume that inhabitants in happier countries tend to live longer, are provided with better health service and can allow more themselves in financial terms.

The independent t-test was performed to compare the adjusted means of people savings and happiness scores. There was a significant difference between means of two variables groups ( $t(97)=4.92$ ,  $p < 0.001$ ). The mean savings were 474 higher than mean of happiness score (95% CI: 350, 824). Therefore, we conclude that the correlation is statistically significant. It is important to mention that correlation between inhabitants' bank account and happiness score is 0.8 in European countries, which is higher than in the world in general. We can suggest that even though in Europe most of the countries are left with positive balance at the end of the month, people feel happier with higher income. However, the correlation between GDP and happiness rating stays on the same level.

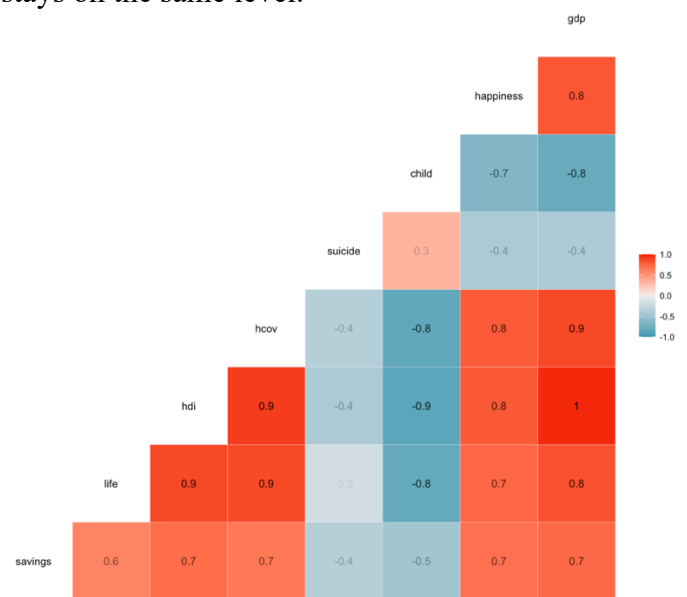


Figure 4

### iv. Clustering

The aim of this subsection is to find similar patterns in relationship between happiness score and financial condition of people in different world regions. For this purpose, k-means clustering algorithm was used, which is based on computing the distance between points in dataset and a centroid and assigning it to cluster. Besides region clustering (Figure 5), European countries as well are presented below (Figure 6).

For region clustering the optimal number of clusters according to analysis of all indices except GAP, Gamma, Gplus and Tau, was 3. Therefore, we can observe that two regions, Australia with Oceania and North America have the highest values of happiness score and potential savings of people. In the middle group we can find countries from Europe, East and Southeast Asia, Middle East regions. It is worth mentioning that the leaders of the happiest countries list are European countries, however, considering also financial condition of people living there, we can notice that North America and Pacific countries are leading.

The last group is represented by South and Central Asia, Africa, South America, Central America and the Caribbean. The last three regions mentioned are almost on the same level with East and Southeast Asia by happiness ranking, however, they are showing lowing results by financial condition.

Let us also look at the clustering plot of European countries. Here we can observe 5 clusters, however, one of them is represented only by Switzerland. Next cluster is represented mostly by Scandinavian countries, however, Finland is included in third cluster, which is quite interesting, since it is the happiest country based on the results of survey. In the same cluster we can also find such countries as: United Kingdom, Luxembourg and Netherlands.

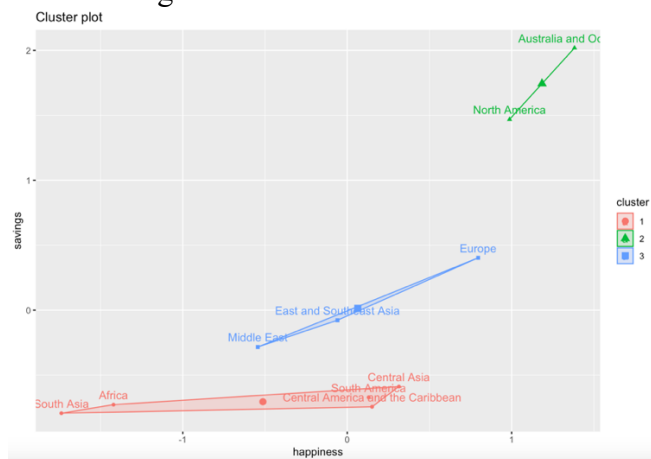


Figure 5

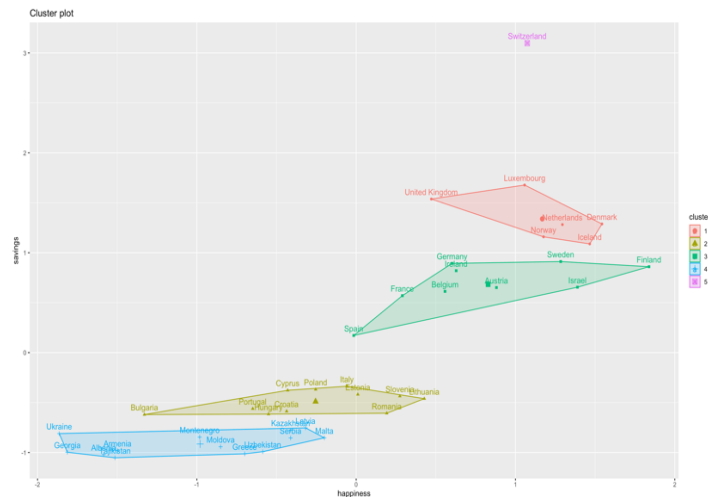


Figure 6

## v. Linear regression model

The results of the following linear regression model aim to predict happiness score based on the amount of money people are left at the end of the month (Figure 7). Looking at the summary closer, we can notice that our model is predicting evenly at both the high and low ends of our dataset, since the median value is 0,2. Coefficients give us opportunity to construct the equation for our model.  $Y = 0.0006680 * x + 5.365$ . As a baseline, if the money at the end of the month equal 0, the happiness score would be 5.365. Then for each 1000\$ of saved money, the happiness score of the country would increase by 0.67 points. In our case standard error value (0.00007128) is larger than coefficient, which shows that coefficient will most likely not be 0. Therefore, savings coefficient is 9.37 standard errors away from zero, which is quite far to say that coefficient will not be 0. P-value is much smaller than 0.05 ( $<2e-16$  and  $3.34e-15$  in our case), which means that our coefficient is significant for our model. According to residual standard error, our actual happiness scores are 0.83 points away from predicted ones on average, which is quite high taking into consideration our actual highest scores (7.8). Savings at the end of the month explain 47.77% of the variation with happiness score. Looking at the F-Statistic and p-value we can conclude that null hypothesis is rejected and there is a relationship between happiness score and amount of money people are left at the end of the month with average expenses.

Now we can look at the provided graphs. Looking at the Normal Q-Q plot we can suggest that the distribution is negatively skewed, which means that our model possibly overestimates happiness

score for countries with smaller income and underestimates for countries with higher one. Scale-Location plot shows us that homoscedasticity is likely to be satisfied for our model. To be sure if homoscedasticity is met, we can perform Breusch-Pagan test, in our case p-value is 0,175 which is higher than 0,05. Therefore, it indicates that we do not have sufficient evidence to claim that heteroscedasticity is present. Based on the Residuals vs Leverage plot, we notice that there are several influential points in our model (two points beneath Cook's distance). However, we can remove those points since the model fits the data. To conclude, taking in consideration all results provided by linear regression model, there is a relationship between happiness score of specific country and amount of money people are left at the end of the month.

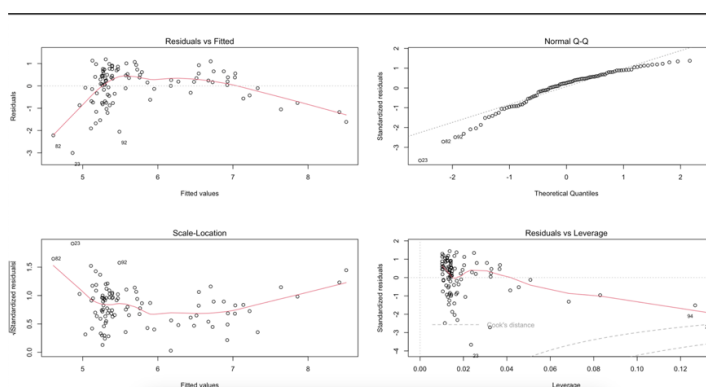


Figure 7

#### 4. Discussion and limitations

The applied statistical methods on the used data provides us with brief overview of the financial condition impact on happiness level around the world. With other features, such as average life expectancy and human development index we were able to observe how different development levels are interrelated with happiness score and cost of living.

One of the biggest limitations of research is that the used data provides the picture only of capital cities, therefore, we cannot make any conclusions regarding the countries overall. On the other hand, it can be considered more as research over capital cities worldwide. We can suggest that in small cities and villages the situation with expenses and income is different from capitals. Since most of the companies vary the prices for products in different countries, but prices inside the country are stable in different cities.

Another limitation would be data quality for African region, since some data provided for cost

of living is absent, as a result, we cannot make fair conclusions regarding this region. In addition, the population could be divided into groups based on age and marital status, since the expenses in families are built differently from people who live alone. In some families there is only one person who is employed, so in such cases income is divided between several people.

Since we did not consider the rent expenses in our study, it is difficult to evaluate if end results are biased. Average property ownership varies from country to country, in some places it is more affordable to rent the property for living than to own one. We should also consider that not everyone lives in equal sized house and there are considerable differences between countries. Property taxation and conditions for purchasing also lies in governmental processes.

We can suggest that people with higher income buy more, therefore, the demand is higher in such countries. For instance, in some developing country the individuals' picture of prosperity would be totally different from the one's in developed country. Several decades ago, people did not have such variety of products and services available, consequently, their demand was lower. Therefore, further research can be made in observing level of consumerism and happiness score worldwide.

#### 5. Conclusion

From the presented analysis we can see that there is relationship between financial condition of inhabitants and happiness rating in different countries of the world. However, we also observed the limitations that were faced during the research.

Despite the obtained results, it is quite difficult to say that happiness depends on the level of income. Happiness is something abstract, nobody can describe it. There is a picture of happiness in society, where an individual is free, has a lot of money, can do whatever s/he wants. In real life our picture of happiness is based on what we consume in terms of information. Information leads our wants, needs and therefore our expectations of life. Nevertheless, information is free to use in various parts of the world, news feed stays different, and the picture of prosperity, values, satisfaction, happiness and misery cannot be compared from country to country. Nevertheless, the life satisfaction is firstly impacted by family environment and then by society.

## 6. Literature

Killingsworth, M. A., Kahneman, D., & Mellers, B. A. (2023). Income and emotional well-being: A conflict resolved. *Proceedings of the National Academy of Sciences of the United States of America*, 120(10). <https://doi.org/10.1073/pnas.2208661120>

Brooks, A. C. (2022b, April 7). How to Increase Happiness, According to Research. *The Atlantic*. <https://www.theatlantic.com/family/archive/2020/04/how-increase-happiness-according-research/609619/>

United Nations. (n.d.). *Human Development Index*. Human Development Reports. <https://hdr.undp.org/data-center/human-development-index#/indicies/HDI>

Kahneman, D., & Deaton, A. (2010b). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, 107(38), 16489–16493. <https://doi.org/10.1073/pnas.1011492107>

Helliwell, J. F. (2023, March 20). *World Happiness Report 2023*. The World Happiness Report. <https://worldhappiness.report/ed/2023/#appendices-and-data>

*Cost of Living*. (n.d.). <https://www.numbeo.com/cost-of-living/>

World Health Organization. <https://www.who.int>

*Kaggle: Your Machine Learning and Data Science Community*. (n.d.). <https://www.kaggle.com/>