

# Happiness Metrics for Global Life Expectancy Analysis

Nicole Chang, Sourish Guntipally, Aaron Park, Brandon To

## BE175 Winter 2022 Final Project Report

### Introduction

In recent years, there has been an increasing importance placed upon individuals' mental health and wellbeing. With growing movements pushing for mental health awareness, many have begun to realize the profound negative impacts of various factors that affect our daily lives such as economy, family, and government. Furthermore, major world events such as the COVID-19 pandemic and more recently the Russia-Ukraine conflict have only accentuated these issues and have proven to be huge contributors weighing on the mental state of the general public and points of concern as society moves forward. Our goal is to evaluate the correlation between happiness and its contributors to life expectancy through the use of machine learning and resampling methods such as Partial Least Squares Regression (PLSR) and Cross Validation to determine and verify the presence of a relationship between the two aforementioned factors.

### Problem Definition

From Kaggle, we chose two different data sets—the World Happiness Report (WHR) [2] and Human Life Expectancy (HLE) [1]. The WHR highlights some of the metrics involved in happiness for countries that participated in the Gallup World Poll, and the average life expectancies of countries each year from 1990-2015. The Gallup World Poll utilizes the Cantril ladder, which asks participants of each country to rank their own lives based on a ladder, with a 10 corresponding to the best possible life and a 0 with the worst possible life. The sample sizes varied, but they were usually less than 3,000 and typically averaged around 1,000 for each country [3]. Samples of these data sets are shown below:

#### *WHR 2015 (Unprocessed) [2]*

	Country	Region	Overall Rank	Happiness Score	Standard Error	Economy	Social support	Healthy life expectancy	Freedom to make life choices	Perceptions of corruption	Generosity	Dystopia Residual
0	Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.34951	0.94143	0.66557	0.41978	0.29678	2.51738
1	Iceland	Western Europe	2	7.561	0.04884	1.30232	1.40223	0.94784	0.62877	0.14145	0.43630	2.70201
2	Denmark	Western Europe	3	7.527	0.03328	1.32548	1.36058	0.87464	0.64938	0.48357	0.34139	2.49204
3	Norway	Western Europe	4	7.522	0.03880	1.45900	1.33095	0.88521	0.66973	0.36503	0.34699	2.46531
4	Canada	North America	5	7.427	0.03553	1.32629	1.32261	0.90563	0.63297	0.32957	0.45811	2.45176

The WHR consists of the used metrics and their definitions as defined by the WHR based on the Cantril Ladder [3] (only main metrics are included; ignored Standard Error, region):

1. **Country**
2. **Overall Rank:** Rank on the world happiness report
3. **Happiness Score:** How happy individuals are
4. **Economy:** GDP per capita
5. **Social support:** A community of family, friends, and acquaintances that provide help in times of need
6. **Healthy Life Expectancy:** Average life in good health
7. **Freedom to make life choices:** One's opportunity to perform an action without being constrained
8. **Perception of corruption:** Trust in the government
9. **Generosity:** Proportion of country that reported donating money to charity
10. **Dystopia Residual:** An imaginary country with the world's least happiest people. Serves as a benchmark to be compared to other countries; where no country performs worse than Dystopia

### *HLE (Unprocessed) [1]*

	Country	Country_Code	Level	Region	1990	1991	1992	1993	1994	1995	...	2010	2011	2012	2013	2014	2015	2016	2017
0	Afghanistan	AFG	National	Total	50.3	51.0	51.6	52.3	52.8	53.4	...	61.0	61.6	62.1	62.5	63.0	63.4	63.8	64.1
1	Afghanistan	AFG	subnational	Central (Kabul Wardak Kapisa Logar Parwan Panj...	51.4	52.12	52.73	53.45	53.96	54.57	...	62.34	63.06	63.67	64.18	64.8	65.3	65.72	66.02
2	Afghanistan	AFG	subnational	Central Highlands (Bamyan Daikundi)	48.59	49.26	49.84	50.52	51.0	51.58	...	58.92	60.04	61.06	61.96	62.96	63.84	64.25	64.55
3	Afghanistan	AFG	subnational	East (Nangarhar Kunar Laghman Nooristan)	54.11	54.86	55.5	56.26	56.8	57.44	...	65.62	65.16	64.61	63.98	63.47	62.88	63.27	63.57

HLE consists of a human's life expectancy per country [1]:

### 1. Human Life Expectancy

A couple of questions that we hope to answer is:

1. How well do the metrics (features) that predict happiness also predict life expectancy?
2. How well does happiness correlate with life expectancy?
3. How accurate are our models at making predictions for new data?
4. How much covariance do our models capture?

## Methods

In order to model our datasets of the World Happiness Report (WHR) [2] and Human Life Expectancy (HLE) [1], the algorithm of Partial Least Squares Regression (PLSR) was implemented. The method of PLSR is commonly used to predict data by maximizing covariance and allows easy interpretation using scores and loadings plots. We used PLSR to predict how the metrics used for the World Happiness Report affect the happiness index scores and life expectancies of countries around the world.

By the end of the data processing and sorting, the metrics we had left for the WHR data set were: Happiness Score, Economy, Social support, Freedom to make life choices, Perceptions of corruption, and Generosity. HLE remained as life expectancy.

Using the processed data, we were able to utilize PLSR and generate scores and loading plots for each year as well as the average values across all 5 years. The generated scores plots showed each country and their correlations with PC1 and PC2. The generated loadings plots showed the weighting of each metric of the WHR in relation to the life expectancy and the happiness score. Within the PLSR modeling, we also calculated the  $R^2Y$ , or covariance for each year by using the score of the dependent variables and independent variables. This helped show how many principal components it would take to explain the covariance between the metrics and happiness scores/life expectancy. Source code for this is found in the second and following sections in *Happiness-Metrics-for-Global-Life-Expectancy-Predictive-Analysis.ipynb*. The PLSR modeling took up a large portion of the source code.

Following the PLSR modeling was the cross validations where we implemented KFold and LeaveOneGroupOut from scikitlearn. This was done to validate our models and observe if they can actually predict new data accurately. To do this, we tested each model using 2 types of cross-validation. For leave one out cross validation using k-folds, we performed resampling by dividing our sample into k number of groups ( $k = 5$ ). As for the LeaveOneGroupOut method, we decided to use this resampling method as a more overarching procedure to see the effects on our prediction model if we dropped WHR factors instead of individual observations to see how omission of metrics would affect the validity of our models.

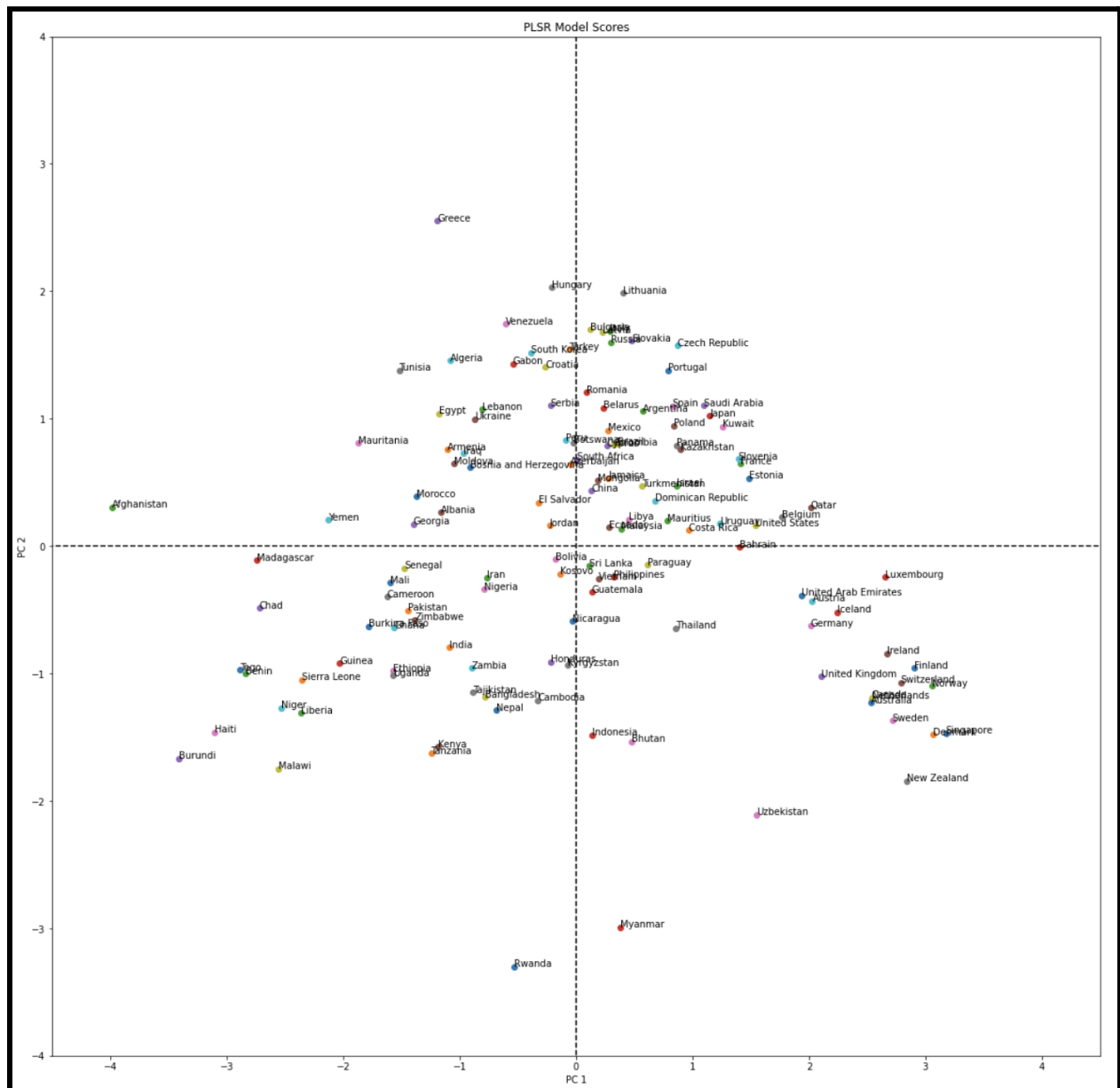
## Results:

### *PLSR Models*

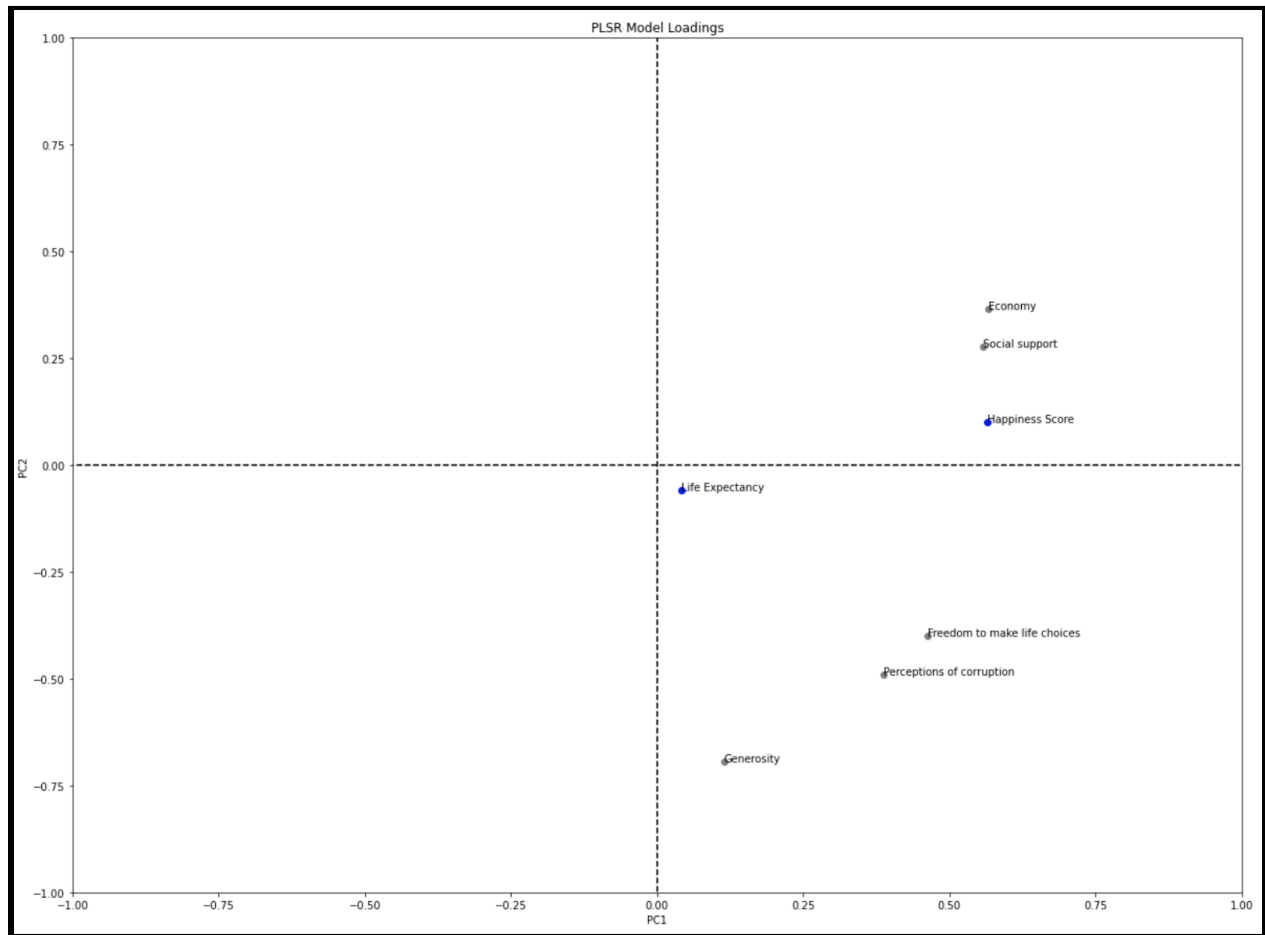
Upon analyzing our results from the generated scores and loadings plots from each PLSR model for the years 2015-2019, we made the general conclusion that there was little to no correlation between happiness and life expectancy. In other words, being happier doesn't necessarily mean you live longer. For example, Japan has one of the highest life expectancies, but often ranks in the 50s or 60s of the WHR for each year. On the other hand, the United States ranks somewhere in the top 15% of countries and has a lower life expectancy. Based on these

findings as well as general common sense, it can be concluded that there are many other factors than just the metrics we examined that influence life expectancy and happiness. A few examples of such factors can include, but are not limited to, geographical environments/locations, climate change, public health, and mental health. An example scores and loadings plot is shown below for the year of 2019.

*PLSR Scores Plot for 2019*

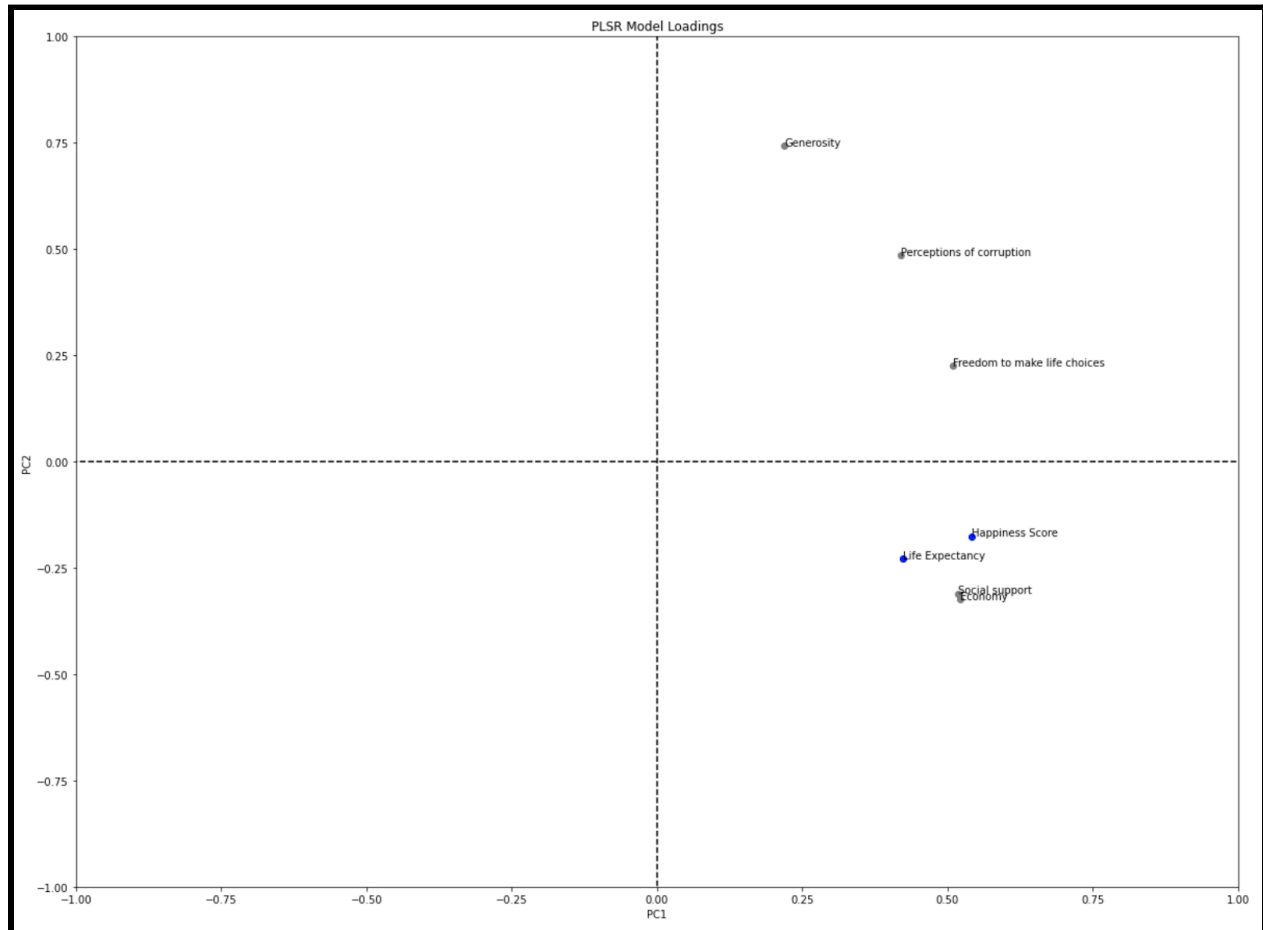


### *PLSR Loadings Plot for 2019*



However, upon performing PLSR modeling for the averaged data across all 5 years, the loadings plot was vastly different from the aforementioned models. As seen in the figure below, life expectancy is more positively correlated along PC1 indicating that happiness does play an integral role in increasing life expectancy. Although, the metrics remained pretty similar to the PLSR models for the 5 years as it was mostly what we expected since economy and social support had the most effect on both life expectancy and happiness score.

### *PLSR Loadings Plot for Averaged Data Over 5 Years*



Unfortunately, while we tried to determine if certain world events had major impacts on particular countries, we determined that it is inconclusive based on the provided data. This is because the data was based on year-to-year, and it is difficult to ascertain if world events had an impact on life expectancy and happiness is not on a monthly basis.

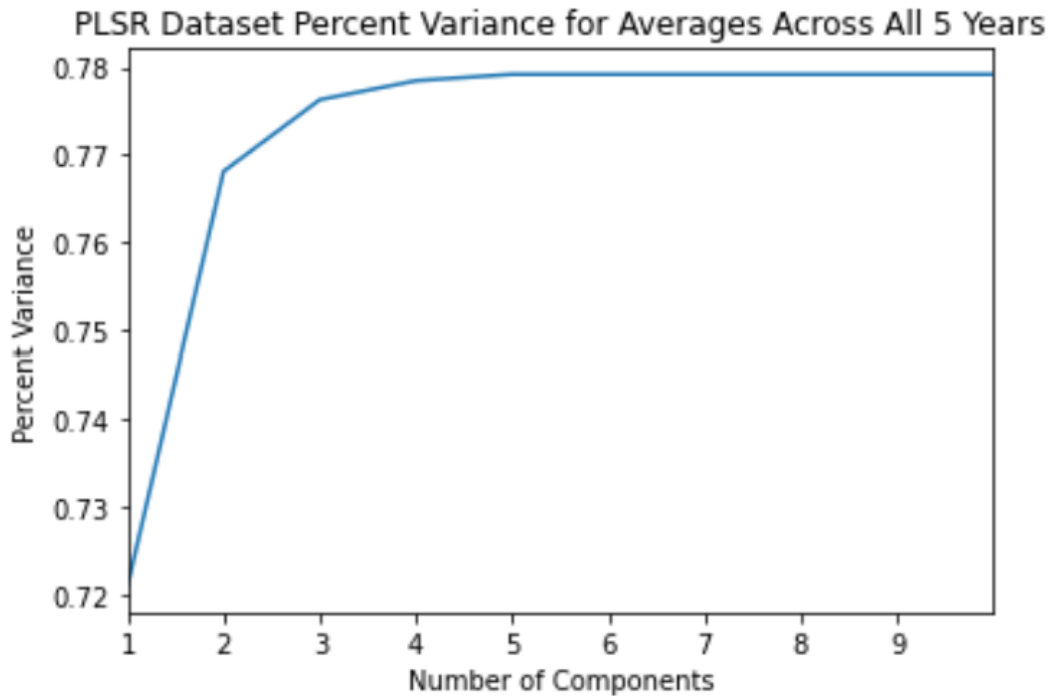
Overall, based solely on the 5 PLSR models for the 5 years, it can be concluded that happiness doesn't correlate with life expectancy. However, based on the averaged PLSR model, it can be concluded that happiness has a significant correlation with life expectancy, meaning that increasing the same metrics that impact happiness index scores can increase life expectancy. This was a surprising finding as our results were dependent on whether our data was averaged over time or taken from a single year.

*Accuracy Values of PLSR Models from K-Fold Cross-Validation*

	Happiness Index Score	Life Expectancy
2015	0.749	-0.154
2016	0.754	-0.114
2017	0.761	-0.147
2018	0.740	-0.027
2019	0.720	-0.188
Averaged Values	0.934	0.559

Based on the table above, our PLSR models were good enough to predict happiness index scores on a per year basis, but their performance was especially good when the data was averaged over 5 years. In general, our PLSR models were extremely poor at predicting future life expectancy values but their performance was improved from nearly 0% accuracy to 55.9% when life expectancy values were averaged. Leave one group out cross-validation had similar results to k-fold CV.

### *Covariance Explained by PLSR*



As observed above in the  $R^2Y$  plot for the averaged data over 5 years, most of the covariance contained within our averaged dataset could be explained with 3 PCs. The covariance plateaus at around 0.78, indicating a high amount of variance existing between our X and Y matrices during PLSR conduction. Alternatively, there was a high correlation between X and Y.



## Citations

[1] Contractor, D. (2022, January 4). Human life expectancy around the world. Kaggle. Retrieved February 8, 2022, from

<https://www.kaggle.com/deepcontractor/human-life-expectancy-around-the-world>

[2] Network, S. D. S. (2019, November 27). World happiness report. Kaggle. Retrieved February 7, 2022, from <https://www.kaggle.com/unsdsn/world-happiness>

[3] World Happiness Report. (n.d.). Retrieved March 8, 2022, from <https://worldhappiness.report/>

[Github Repository for the project](#)