[redacted] Damary Gutierrez Hernandez [redacted]

COMPSCI216 - Everything Data
28 April 2021

INTRODUCTION

The causes and effects of happiness have been topics of discourse amongst economists, sociologists, and psychologists for the past several decades. The study of happiness is one of intense research because of its implications to different fields such as health, economics, and politics. In fact, life satisfaction and happiness have been shown to broaden one's perspective, encourage political participation, and facilitate interpersonal relationships (Veenhoven, 1988). Happiness also affects life expectancy and overall physical and mental well-being in healthy populations (Veenhoven, 2008; Post, 2005; Sabitini, 2015). In addition, happiness in the workplace can also result in an approximately 12% increase in productivity (Oswald, 2015). With so many positive outcomes that come from happiness, it is important to understand what factors play a role in promoting it.

Previous studies have shown that happiness is very subjective and can be largely dependent on cultural constructions (Uchida, 2004). Although there are cultural differences in how happiness is defined, demographic differences including race, age, sex, and education only account for small variation in happiness measures (Diener *et al.*, 1999). In addition, happiness across different cultures can still be attributed to key factors including self-esteem, sense of control, optimism, and social capital (Myers and Diener, 1995; Leung *et al.*, 2011). The World Happiness Report is a survey to understand the state of global happiness and what factors can be attributed to a country's subjective well-being (Helliwell *et al.*, 2020). In previous years, the World Happiness Report has analyzed general trends regarding global happiness and also how natural disasters or economic turmoil plays a role in perceived happiness.

The COVID-19 pandemic has had a large effect on subjective happiness across the world. Lockdown and public health measures to stop the spread of disease, financial loss and worsening economies, and political turmoil exacerbated by the pandemic have had a toll on the mental well-being of people (Pfefferbaum and North, 2020). In addition, the COVID-19 pandemic has revealed other factors that may play a role in happiness and well-being. Here we explore the question: what is the relationship between happiness and a country's COVID-19 response? This study also seeks to understand the factors most important to groups of countries with different happiness scores or demographics.

The relevance of this question cannot be understated. Understanding happiness has long been a goal of the scientific community, for it has substantial real-world implications for individuals and entire communities. In discovering how happiness is related to a country's COVID-19 response, we can better predict the impacts that future pandemics will have on nations. We can also determine the ways that a country's disease response is a factor in determining the citizenry's happiness.

SUMMARY OF RESULTS

In this study, we used linear and nonlinear regression analysis to feasibly show correlations between COVID-19 vaccination process, case rates by country, and the world happiness report data. When looking at the relationship between COVID-19 vaccine progress and overall pre-pandemic happiness factors, we show that there are positive correlations between indicators of strong economies, social capital, and vaccine rollout. Specifically, there is a positive correlation between GDP per capita, social support, generosity, and vaccine progress. Interestingly, perceptions of corruption and vaccine progress were also positively correlated. When analyzing the relationship between COVID-19 cases and happiness factors, we show that GDP per capita, social support, healthy life expectancy at birth, and life ladder had the strongest correlation with COVID-19 case counts in April 2021. Using this data, we also used machine learning tools to predict vaccination efforts from happiness data and vice versa. We show that there is notenough data to confidently predict vaccine rollout from happiness data as well as happiness from vaccine rollout data.

Since previous work has shown that demographic differences account for small variation in overall happiness (Diener *et al.*, 1999). Therefore, this study attempted to understand how differences in overall happiness can be attributed to demographic differences and if that affects COVID-19 case counts. Through the use of linear and nonlinear regression analysis, we have observed a marked difference in a country's COVID-19 response based on demographic factors such as when countries are grouped based on similar happiness scores or age demographics.

DATA SOURCES

In order to tackle our question, we use two different datasets with appropriate measures (Table 1). The first dataset (sourced by Our World in Data) contains information primarily representing COVID-19 testing, vaccinations, and cases (https://github.com/owid/covid-19-data/tree/master/public/data). This dataset also includes demographic data related to age. We also use data from the World Happiness Report that contains happiness measures for each country from 2008-2020 (https://worldhappiness.report/ed/2021/#appendices-and-data). Our code can be accessed in different branches at https://github.com/amrbedawi/216_Project.git. Supplemental materials can be accessed at this link.

Table 1. Datasets and relevant measures used in this study.

Dataset	**Measures**
Our World in Data COVID-19 (April, 2021)	People vaccinated per hundred; Daily vaccinations per million; Total COVID-19 cases; Median age; Proportion of population 65 and older; Population Density
World Happiness Report (March, 2021)	Year; Life Ladder; Log GDP per capita; Social support; Healthy life expectancy at birth; Freedom to make life choices; Generosity; Perceptions of corruption; Positive affect; Negative affect

RESULTS AND METHODS

1. How are the happiness scores of countries related to their COVID-19 vaccination progress?

There were two metrics we used to relate COVID-19 vaccination progress with the happiness scores. The first was vaccinations per hundred, supplied by the COVID-19 World Vaccinations Progress dataset. This is the total number of fully vaccinated people against COVID-19 per hundred given by countries on or before April 11, 2021. Countries that did not report a total vaccinations figure between April 1-11, 2021 were excluded. The second metric was the seven-day average daily vaccination rate per 100 (i.e, how many COVID-19 vaccines were administered in the country per day per million people) from April 1-7, 2021. Countries that did not report any data in this time frame were excluded. The reason a seven day average was chosen is that the daily average can vary dramatically within a single week, so this should smooth out the data for each country to give a more precise metric. These two metrics—total administered COVID-19 vaccines and current vaccination rate (as of April 2021)—can be used to numerically evaluate a country's progress in their individual COVID-19 vaccination campaigns.

Table 2: Summary of regression results

	Linear		**Quadratic**	
	R^2	Adjusted R^2	R^2	Adjusted R^2
Vaccination rate All parameters	.322	.251	.406	.266
Vaccination rate Selected parameters	.311	.281	.383	.325
Total vaccinations All parameters	.426	.363	.535	.420
Total vaccinations Selected parameters	.418	.385	.497	.435

The World Happiness Report dataset has several data points for each country, and we used the 2019 data (released in 2020) to do a linear and nonlinear regression analysis to analyze how the pre-pandemic happiness scores of a country relate to their total vaccinations and vaccination rate. 92 countries reported enough data to include in the model for total vaccinations (number of fully vaccinated people), and 95 countries reported enough data to include in the model for vaccination rate (administered vaccines per day). First, we did a linear regression using the nine happiness data points for each country as the data used to predict the two vaccine metrics. The coefficients and p-values can be seen in Table A1 (in the appendix in supplemental materials). Then, we did a nonlinear regression by adding in quadratic terms. The coefficients for the quadratic models are in Table A2. Finally, we repeated both the linear and quadratic models with the removal of some selected parameters that had high p-values. This was done to make the model more simple, to avoid overfitting on uncorrelated or unhelpful data, and to avoid collinearities. The slimmed down models are in tables A1 and A2. A summary table showing only the R^2 values can be found in Table 2.

There are two main takeaways regarding the accuracy of the models. First, our method of narrowing down the parameters to those with small p-values did not significantly decrease the R^2 value of the models. In fact, the adjusted R^2 values increased for both the linear and quadratic models for vaccination rate and total vaccinations. The adjusted R^2 value takes into account the number of parameters used, and is higher if the variables included increase the accuracy of the model by more than what would occur by chance. This indicates that the models with the fewer selected parameters work just as well than when including all the happiness data. The p-values for each individual parameter on the slimmed down models are on average much smaller than the full models, indicating each individual parameter in the slimmed down models were more significant. Comparing the plots of figures A1 and A2 in the appendix shows that the model doesn't change very much when dropping the selected parameters.

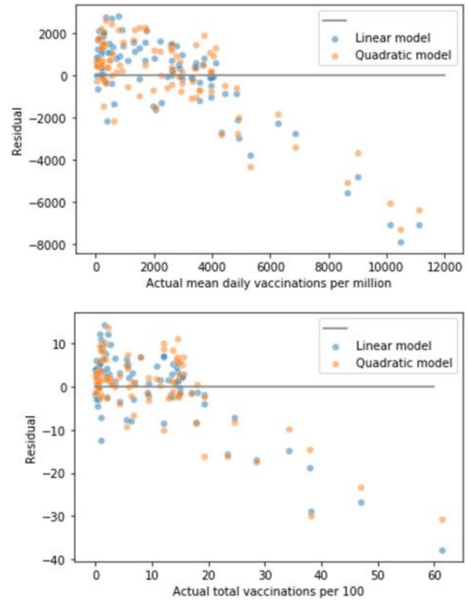

Figure 1: Residual plots for regression models with selected parameters

Second, using a nonlinear model increased the accuracy when mapping the world happiness data to COVID-19 vaccinations. This is seen in Figure 1, showing the residuals for the slimmed down models. There was a trend in the residuals that was seen in the linear models, motivating the use of nonlinear terms. It can be seen in the right tail in the residuals in Figure 1 that using quadratic terms decreased the error, albeit not completely. However, the adjusted R^2 values for the quadratic models were greater than their linear counterparts, showing that the nonlinear model does a better job mapping the happiness data to vaccination rates without overfitting by using too many parameters. There is still a trend in the residuals indicating that the model may benefit by other nonlinear terms. It is also possible that there is another variable that is not in the dataset that explains why these outlier countries with much higher vaccination rates have underestimated outputs.

It is interesting to investigate which happiness factors were positively and negatively correlated with vaccination progress. We will use the slimmed down linear models to get a sense for which parameters were associated with high and low vaccination progress. For total vaccinations: life ladder (a metric based on surveys about contentment with life), log GDP per capita, generosity, and perceptions of corruption were positively correlated. Positive affect (a metric based on surveys about recent positive feelings) was negatively correlated with total vaccinations. All p-values were less than 0.13, but most were smaller than this (Table A1). It makes sense that many of these parameters are associated with higher vaccination progress (such as wealthier countries doing better perhaps due to better infrastructure and vaccine purchasing power). It is interesting that perceptions of corruption were positively associated, but more research is needed to explain why. It is important to note that this is a measure of perceptions of corruption among people, not if there literally is corruption. For vaccination rate, the parameters that were positively associated were life ladder, log GDP per capita, social support, and perceptions of corruption. However, only GDP in the linear model had a small p-value (<0.05). No parameters were negatively associated. The quadratic model may explain more, since the linear model only confidently could suggest a link between GDP and vaccination progress. Using the squared parameters, it is seen that these four parameters are positively associated with small p-values less than 0.06. Again, interestingly perceptions of corruption were positively associated with vaccination progress.

It is important to note that correlation does not imply causation; while these correlations were found in the data, strong research and evidence is required to explain why the correlations exist, and if they are causal relationships. A possible explanation for the positive associations with GDP is to assume that wealthier countries have higher purchasing power for vaccines and better infrastructure, so they may vaccinate faster. However, the reasons why perceptions of corruption were positively correlated, and positive affect negatively correlated. are not clear, and more research is needed to answer these questions.

2. Can we predict vaccination efforts based on happiness data and predict happiness based on vaccination data?

In our exploration of the relationship between happiness and vaccination rates, we looked to see how well we could predict vaccination efforts based on the factors that the World Happiness Report deems important to happiness and how well our predictions would do in the reverse direction. We define vaccination efforts as the number of people who received at least one vaccine dose per one hundred and the number of fully vaccinated people per one hundred people. Happiness is determined by 14 different factors including GDP, life expectancy, and social support. For our predictions, we utilized the database containing the world happiness report for 2020 and a covid vaccination database. We merged the two databases and dropped any countries that did not have happiness data or vaccination data. Because some of the features were on a much larger scale than the rest, we had to transform the data using the StandardScalar tool from sklearn.

We split the data and kept two-thirds as training data and the rest as testing data. Table 3 describes the performance of the models we used. First we attempted to predict vaccination efforts based on happiness factors using a linear regression model. This model was not fruitful. On the testing data, our model had a mean squared error of 520.6 and an R^2 value of $3.5x10^{-5}$. We then utilized a K Nearest Neighbors Regressor model. After performing grid search, we found that the optimal hyperparameter was 2 neighbors. This did not perform well either. On the testing data, this model had a mean squared error of 466.3 and an R^2 value of .085. Our final model was a three-layer neural network regressor that utilized a rectified linear unit activation function. This model also performed poorly, as it had a mean squared error of 481.5 and an R^2 value of .042.

Table 3. Performance of machine learning models to predict vaccination efforts from happiness data.

	Linear Regression	**K Nearest Neighbor**	**Neural Network**
Mean Squared Error	520.6	466.3	481.5
R^2 value	$3.5x10^{-5}$	0.085	0.042

Using the same methods as above, we then attempted to predict factors of happiness based on vaccination efforts. Table 4 describes the performance of the models we used. The linear regression model performed poorly with a mean squared error of 1.5 and an R^2 value of -0.53. The K Nearest Neighbors Regressor performed significantly better at predicting happiness factors. With an optimal hyperparameter of 18 neighbors, it had a mean squared error of 0.596 and an R^2 value of 0.37. The neural network had a low mean squared error of .9, but the R^2 value was also low at .027.

Table 4. Performance of machine learning models to predict happiness factors from vaccination data.

	Linear Regression	**K Nearest Neighbor**	**Neural Network**
Mean Squared Error	1.5	0.596	0.9
r^2 value	-0.53	0.37	0.027

For every model we created, the model performed significantly better when predicting happiness factors from vaccination data than they did predicting vaccination efforts from happiness data. These results suggest that it is easier to predict happiness factors from vaccination efforts than it is to predict vaccination efforts based on factors of happiness. The relationship does not seem to behave in a linear fashion, as the linear models do not do well on the data, while the K Nearest Neighbors Regression models do better. These machine learning models would benefit greatly from more data. Covid vaccination data is new and relatively sparse, and the world happiness report only takes into account a couple of countries. Our work only utilizes 2020 data, and I predict that including information from more years would increase our predictive capabilities. Additionally, more sophisticated models like convolutional neural networks or support vector machines could possibly produce better results.

3. How are the happiness scores of countries related to their COVID-19 case rate data?

A similar approach was applied to analyze the relationship between the happiness variables of the countries (i.e., life ladder, social support, log GDP per capita, freedom to make life choices, perception of corruption, generosity, healthy life expectancy at birth, positive affect, and negative affect) and total positive COVID-19 tests in that country as of April 2021. The number of COVID-19 cases were summed together to acquire total number of cases for each separate country. The dataset was then merged based on country name. As shown below, the nine happiness variables were each separately plotted against total cases. Linear regression models were then separately created for each pairing (Figure 2).

From the comparisons, no R^2 value was above 0.5. The highest value was with the comparison of log GDP per Capita with total COVID cases which had an R^2 value of 0.310. The perception of corruption and negative affect both individually had an R^2 value of 0.001. The overall results highlight the lack of a direct linear relationship between any of the individual happiness variables with total positive cases.

Exploring the relationship of happiness and total covid cases in each country further, we applied a linear regression with all combined happiness variables of the country compared to the total COVID cases (Table 5). From the regression, we were able to increase the R^2 value to 0.555 with an MSE of 431674767.6. This increase would suggest a more integrated response between the happiness variables and total covid cases. We additionally ran a nonlinear quadratic regression model comparing all combined happiness variables to the total COVID cases. From such a model, we recorded an R^2 value of 0.867 and an MSE of 128495541.5. This high R^2 value further supports the idea that such a response is reliant on the combination of many happiness variables. Additionally, the higher R^2 value suggests a quadratic relationship. However, there is a concern of overfitting.

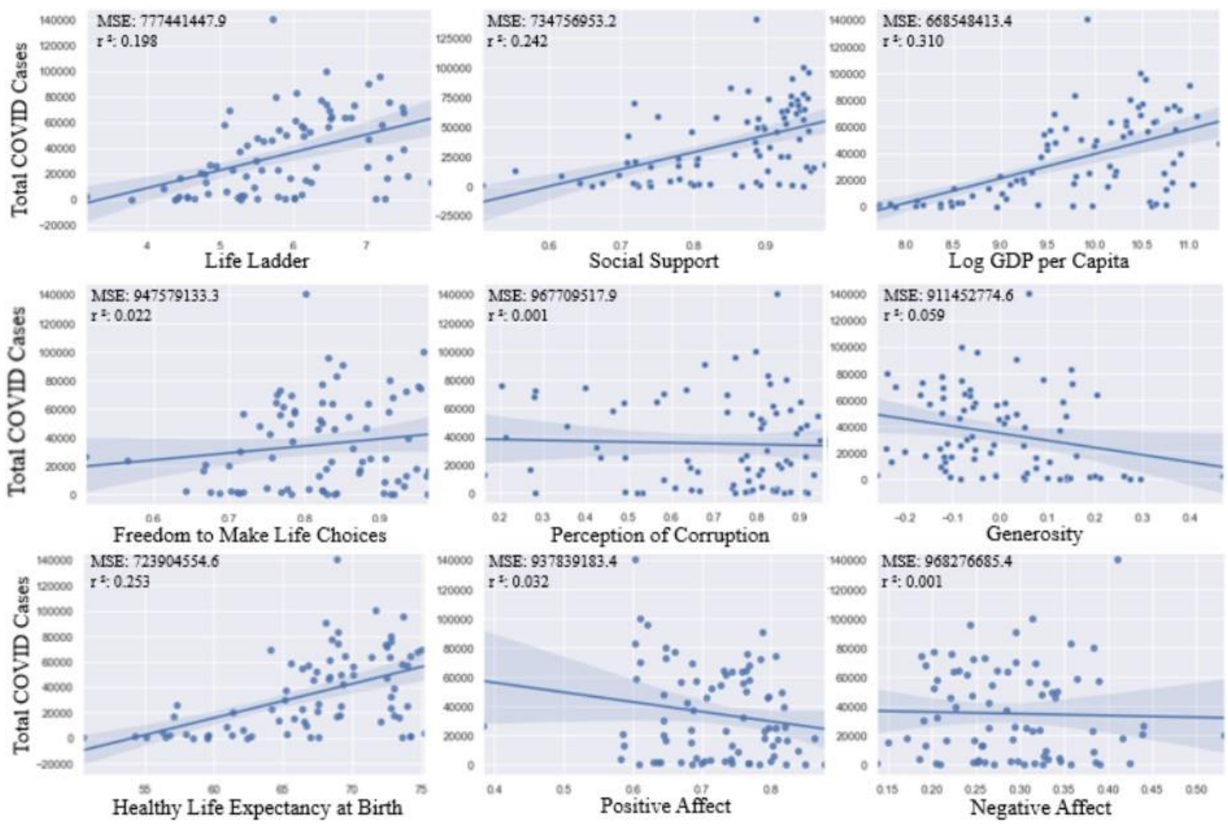

Figure 2. Linear regression models of total COVID cases compared to separate happiness variables. We observed a low relationship among each of the separate happiness variables. The highest R^2 value recorded occurred with the Log GDP per Capita comparison. The lowest R^2 values occurred with Perceptions of Corruption and Negative Affect.

Table 5. Regression Analyses for combined happiness variables and total COVID cases. The nonlinear quadratic regression had a higher R^2 compared to the linear regression model.

	Linear Regression of Combined Happiness Variables	Nonlinear Quadratic Regression of Combined Happiness Variables
Mean Squared Error	431674767.6095129	128495541.54626313
R^2 value	0.5545592089296316	0.8674067608993102

4. Between countries with similar happiness scores or similar demographics, how do demographic factors play a role in their COVID-19 response?

Before choosing countries with similar happiness scores, we first created a linear regression model in order to look at the general relationship between demographic factors such as the median age (“median_age”), the percent of the population aged 65 years or older (“aged_65_older”), the population density (“population_density”), and the GDP per capita (“gdp_per_capita”) of each country and the COVID-19 response of those countries as measured by total confirmed cases of the coronavirus per million people (“total_cases_per_million”). For this analysis, countries without values in each of the demographic columns were dropped, leaving us with a dataset of 140 countries total. Table 6 summarizes the coefficients which describe the relationship between the selected demographic factors and a country’s cases.

Table 6. Linear model coefficients between selected demographic factors and COVID-19 cases for all countries.

Coefficient	median_age	aged_65_older	population_density	gdp_per_capita
Predicted total_cases_per_million for all countries	1.31×10^{3}	6.88×10^{2}	-4.78	2.93×10^{-1}

To determine how well this model does at predicting a country’s total cases per million given the demographic factors, we calculated the model’s mean squared error (MSE) and R squared (R^2) values. For this model, MSE = 533062441.598 and R^2 = 0.432. Since our MSE is extremely large and our R^2 value is not very close to 1, we can conclude that this model is not very good at predicting a country’s total cases per million.

When comparing our model with the real data, we can see that our desired relationship between the variables is most likely not linear. Figure 3 provides visualizations that plot the various independent variables against our dependent variable.

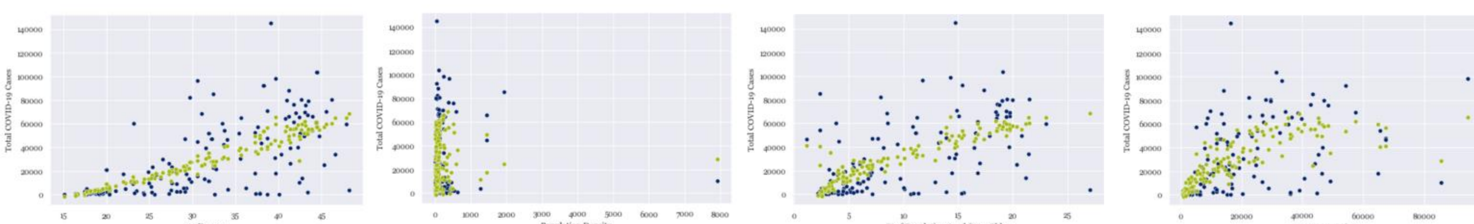

Figure 3. Linear model predictions for all countries (in green) vs. real values (in blue) for each demographic factor.

A nonlinear relationship is most prevalent in the following variables: “median_age”, “aged_65_older”, and “gdp_per_capita”. As such, it would be best to transform said variables to arrive at a better-fitting model. The better the model, the more secure we can be in comparing the coefficients of the COVID-19 response of countries with similar happiness scores or demographic factors to those of this baseline model.

After creating a baseline model, we then sorted countries by happiness scores. The countries with the highest five happiness scores were sorted into their own dataset. The same was done with the countries with the lowest five happiness scores and with countries that we determined had similar demographics. Thus, the top five scores consisted of Norway, Iceland, Switzerland, Denmark, and

Finland, the bottom five scores consisted of Afghanistan, South Sudan, Zimbabwe, Rwanda, and the Central African Republic, and the countries with similar demographics consisted of the United States, South Africa, and the United Kingdom.

After creating a linear regression model for each of the subgroups, we arrive at the following coefficients in Table 7.

Table 7. Linear model coefficients between selected demographic factors and COVID-19 cases for the five countries with the highest happiness scores, the five countries with the lowest happiness scores, and the three countries with similar demographics.

Coefficient	**median_age**	**aged_65_older**	**population_density**	**gdp_per_capita**
Predicted total_cases_per_million for top five countries	1.05×10^{4}	-9.68×10^{3}	1.49×10^{2}	-2.34×10^{-2}
Predicted total_cases_per_million for bottom five countries	8.74×10^{2}	2.93×10^{2}	-3.17	5.09×10^{-1}
Predicted total_cases_per_million for similar countries	-0.58	-0.60	-21.59	1.56

As a result of the small number of countries included within each model, the MSE value is very small and the R^2 values are 1.

After applying multiple transformations and choosing the one that minimized MSE and maximized R^2, we arrived at a cubic model with an MSE of 294820857.175 and an R^2 of 0.686. While the MSE is still very high, our R^2 value is much closer to 1. We then arrived at the following coefficients seen in Table 8.

Table 8. Cubic model coefficients between selected demographic factors and COVID-19 cases for all countries. The 22 interaction terms were removed for clarity.

Predicted total_cases_per_million for all countries	**median_age**	**aged_65_older**	**population_density**	**gdp_per_capita**
$(X)^1$	-4.29×10^{4}	1.97×10^{1}	-1.93×10^{2}	1.39×10^{1}
$(X)^2$	2.09×10^{3}	-4.14×10^{2}	3.05×10^{1}	-1.16
$(X)^3$	-6.31×10^{2}	-4.51×10^{1}	1.76×10^{-1}	5.42×10^{-2}

In an effort to compare how our new cubic model performs with respect to our real data, we have provided visualizations in Figure 4 that plot the various independent variables against our dependent variable.

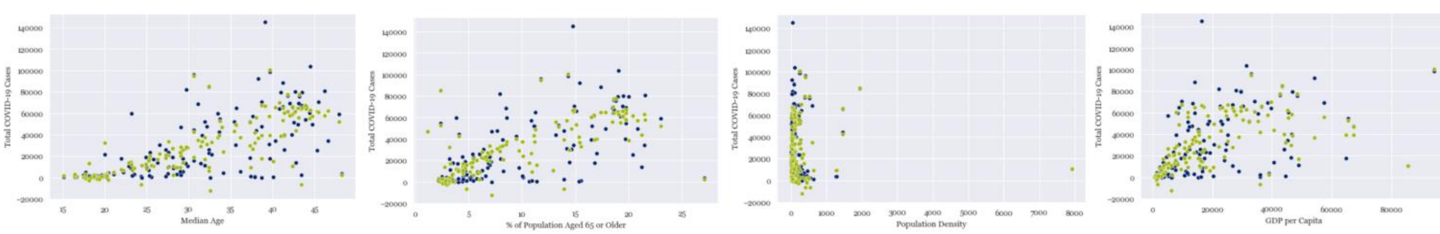

Figure 4. Cubic model predictions for all countries (in green) vs. real values (in blue) for each demographic factor.

As we can see through the visualizations, our cubic model better follows the real data. As such, we can be much more confident in comparing the coefficients between our new baseline model and our subgroup models.

After transforming our model for each of the subgroups, we also have the following coefficients seen in Table 9.

Table 9. Cubic model coefficients between selected demographic factors and COVID-19 cases for the five countries with the highest happiness scores, the five countries with the lowest happiness scores, and the three countries with similar demographics. The 22 interaction terms for each model were removed for clarity. For each of these models, the MSE ≈ 0 and $R^2 \approx 1$.

Predicted total_cases_per_million for top five countries	**median_age**	**aged_65_older**	**population_density**	**gdp_per_capita**
$(X)^1$	-2.27×10^{-14}	-1.98×10^{-14}	1.16×10^{-12}	2.94×10^{-11}
$(X)^2$	-1.94×10^{-12}	-1.28×10^{-12}	4.69×10^{-11}	4.01×10^{-11}
$(X)^3$	-8.36×10^{-13}	2.53×10^{-11}	-1.15×10^{-10}	3.21×10^{-12}
Predicted total_cases_per_million for bottom five countries	**...**	**...**	**...**	**...**
$(X)^1$	5.56×10^{-12}	2.48×10^{-12}	-3.10×10^{-11}	-5.98×10^{-10}
$(X)^2$	2.13×10^{-10}	6.16×10^{-11}	-1.89×10^{-10}	-1.99×10^{-10}
$(X)^3$	1.43×10^{-11}	2.05×10^{-11}	1.83×10^{-9}	3.50×10^{-8}
Predicted total_cases_per_million for similar countries	**...**	**...**	**...**	**...**
$(X)^1$	7.79×10^{-19}	7.81×10^{-19}	1.94×10^{-17}	9.46×10^{-16}
$(X)^2$	5.39×10^{-17}	3.68×10^{-17}	8.27×10^{-16}	5.16×10^{-14}
$(X)^3$	1.96×10^{-17}	3.94×10^{-16}	3.13×10^{-14}	6.11×10^{-15}

While improvements can be made to the transformed general model (as evidenced by the fact that the R^2 value can still be improved), we can compare the coefficients of the transformed general model (our baseline) with those of the subgroups to see which demographic factors are particularly prone to change. Since our transformed models for each of the subgroups seems to be suffering from some overfitting—especially since our linear models already minimized our MSE and maximized our R^2—it might be more advantageous to compare the linear terms of our cubic model with the linear terms of our non-transformed subgroup models.

The model consisting of countries with similar demographics is the most different from the baseline. This implies that the baseline is better at modeling higher happiness scores than more moderate ones (as in the case of the model consisting of countries with similar demographics). Such an occurrence is most likely due to the fact that the distribution of happiness scores skews left so we have more information regarding higher happiness scores. It is also interesting to note the difference in signs for certain coefficients. For example, our baseline has 1.97×10^1 as a coefficient for “aged_65_older” while the coefficient for the same variable in the subgroup consisting of the countries with the top happiness scores is -9.68×10^3 (a change from positive to negative, respectively). From this, we can infer that when the percentage of the population that is 65 years or older increases, we predict less total cases of COVID-19 for the subgroup model than for the baseline model. Why is this the case? It could be that the healthcare in the countries with the highest happiness scores is accessible enough to mitigate the spread of COVID-19 among the older population or that the policies set by these governments are more effective among the 65+ age group than is generally true among all countries.

It is important to note that in looking only at a subset of our coefficients, we are likely missing the effect that interaction terms have on our model; it could be the case that some of the more impactful terms are hidden within the interaction terms that we have left out. However, we can still tentatively emphasize the importance of taking demographic factors into account when analyzing a country’s COVID-19 response. As seen within the subgroups, the models can vary wildly. For future models, it might be advantageous to perform other transformations in order to increase our R^2 value.

LIMITATIONS AND FUTURE WORK

In all our analyses, we used both linear and nonlinear modeling to investigate the relationship between COVID-19 vaccine rates, case counts, and happiness scores across countries of different demographics. Although we used nonlinear modeling to increase our R^2 value, the quadratic, cubic, and other nonlinear models are not able to fully capture and explain the relationship between COVID-19 outcomes and happiness factors. Therefore, we cannot conclude that there is a strong correlation between these factors.

In addition, correlation does not equal causation. Thus, the weak correlation of our datasets does not suggest that happiness factors cause changes in COVID-19 outcomes and vice versa. This was confirmed with our predictive machine learning models since they were unable to confidently predict happiness scores from vaccination rates as well as vaccination rates from happiness scores. Therefore, our results show links between happiness data and COVID-19 performance measures, but cannot explain them.

In the future, to increase R^2 values for our analyses and improve predictive models, we can use more sophisticated models such as convolutional neural networks or support vector machines. In addition, we could use K Nearest Neighbor Regression Models to determine the optimal hyperparameters for each of our analyses instead of deciding on quadratic or cubic regression models. Finally, obtaining more data would be beneficial in our analyses by using the happiness data from previous years.

The World Happiness Report releases data annually and 2021 data will be released in March 2022. Therefore, it would be interesting to see how happiness scores will change from before to after the COVID-19 pandemic and if COVID-19 case or vaccine data is connected to these changes. In other words, do countries with better COVID-19 responses have greater or smaller changes in happiness?

In addition, our results provide a relationship between COVID-19 data and happiness data but do not provide a strong explanation as to why we see correlations in the data. Thus, it would be interesting to investigate why we see these relationships such as the positive association between COVID-19 vaccination progress and perceptions of corruption. Is it because more corrupt countries may inflate their progress? Is it because more corrupt countries may have less regulatory guidelines for vaccine approval?

Although we have provided preliminary analyses on happiness and COVID-19, there is more to be done to bring light on how COVID-19 affects social capital and happiness.

Bibliography

Diener, Ed, et al. "Subjective Well-Being: Three Decades of Progress." *Psychological Bulletin*, vol. 125, no. 2, 1999, pp. 276–302. *Crossref*, doi:10.1037/0033-2909.125.2.276.

Greyling, Talita, et al. *A Tale of Three Countries: How Did Covid-19 Lockdown Impact Happiness?* p. 19.

Leung, Ambrose, et al. "Searching for Happiness: The Importance of Social Capital." *Journal of Happiness Studies*, vol. 12, no. 3, June 2011, pp. 443–62. *Crossref*, doi:10.1007/s10902-010-9208-8.

Myers, David G., and Ed Diener. "Who Is Happy?" *Psychological Science*, vol. 6, no. 1, Jan. 1995, pp. 10–19. *Crossref*, doi:10.1111/j.1467-9280.1995.tb00298.x.

Oswald, Andrew J., et al. "Happiness and Productivity." *Journal of Labor Economics*, vol. 33, no. 4, Oct. 2015, pp. 789–822. *Crossref*, doi:10.1086/681096.

Pfefferbaum, Betty, and Carol S. North. "Mental Health and the Covid-19 Pandemic." *N Engl j Med*, 2020, p. 3.

Post, Stephen G. "Altruism, Happiness, and Health: It's Good to Be Good." *International Journal of Behavioral Medicine*, vol. 12, no. 2, June 2005, pp. 66–77. *Crossref*, doi:10.1207/s15327558ijbm1202_4.

Sabatini, Fabio. "The Relationship between Happiness and Health: Evidence from Italy." *Social Science & Medicine*, vol. 114, Aug. 2014, pp. 178–87. *Crossref*, doi:10.1016/j.socscimed.2014.05.024.

Uchida, Yukiko, et al. "Cultural Constructions of Happiness: Theory and Emprical Evidence." *Journal of Happiness Studies*, vol. 5, no. 3, 2004, pp. 223–39. *Crossref*, doi:10.1007/s10902-004-8785-9.

Veenhoven, R. "Healthy Happiness: Effects of Happiness on Physical Health and the Consequences for Preventive Health Care." *Journal of Happiness Studies*, vol. 9, no. 3, Sept. 2008, pp. 449–69. *Crossref*, doi:10.1007/s10902-006-9042-1.

Veenhoven, Ruut. *The Utility of Happiness*. p. 22.

Yıldırım, Murat, and Abdurrahim Güler. "Positivity Explains How COVID-19 Perceived Risk Increases Death Distress and Reduces Happiness." *Personality and Individual Differences*, vol. 168, Jan. 2021, p. 110347. *Crossref*, doi:10.1016/j.paid.2020.110347.