How the onset of the COVID-19 Pandemic Affected the Happiness of Different Regions of the World

Mary Nathalie G. Dela Cruz

Abstract-- **The COVID-19 pandemic is a global health crisis that has disrupted the lives of billions of people. This study aims to analyze the effect of the onset of the pandemic on the collective happiness of individuals from different regions in the world based on the 2020-2021 data of the World Happiness Report. The variation in happiness score across regions and over time was examined using one-way ANOVA and paired t-test. Results showed that there was a significant difference in happiness score among regions for both 2020 and 2021, but no significant difference in happiness score due to time, except for Latin America and Caribbean region. This suggests that region is a more important factor than the period of onset of the pandemic when explaining happiness score. The relationship between happiness score and six factors was also examined using Pearson correlation and multiple linear regression. The results indicated that three factors (logged GDP per capita, social support, and healthy life expectancy) had a strong positive correlation with happiness score, while one factor (perceptions of corruption) had a moderate negative correlation and one factor (freedom to make life choices) had a moderate positive correlation. Generosity does not correlate with happiness score. The regression model, which excluded generosity as a non-correlated factor, was statistically significant and explained 75.4% of the variation in happiness score, but violated the normality assumption of errors. A better model that meets the assumption of normality of errors is recommended to be deployed.**

Index Terms — happiness, pandemic, regression, hypothesis testing, ANOVA

I. INTRODUCTION

The health pandemic brought on by the outbreak of the Coronavirus (COVID-19) has multiple negative implications for the world. At the time of writing this paper, over 774 million people have tested positive for the virus, and over 7 million lives have been claimed by the virus [1]. It has also resulted in massive declines in GDP, income, and employment, as well as rising levels of inequality and poverty [2]. These numbers, however, do not capture the full extent of the pandemic's impact. Along with the widespread upheaval of social and economic life comes a higher level of personal psychological distress. Studies on COVID-19 and its link to well-being have shown that during the pandemic, people's happiness has diminished while reporting a growing number of negative emotions [3]. Google searches about boredom, loneliness, worry, and sadness have also significantly increased since [4].

These findings are consistent with the results of the World Happiness Report, which measures happiness by asking how satisfied people are with their lives, based on a scale from 0 to 10. The report correlates happiness scores with various factors, such as GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perception of corruption, and annually ranks countries of the world based on their happiness scores. The 2021 World Happiness Report, to be specific, showcases the effect of the COVID-19 pandemic on the happiness of people through surveys conducted in 2020 and early 2021 [5-6].

This study aims to quantitatively analyze the impact of the onset of COVID-19 on the collective happiness of individuals from different regions in the world based on the 2020-2021 data from the World Happiness Report. A comprehensive analysis is done by exploring the following questions surrounding the onset of the pandemic: How does happiness compare across regions in the world? How has the happiness of different regions changed from the 2020 report to the 2021 report? How do factors such as GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perception of corruption influence the change in happiness? To answer these questions, hypothesis testing, correlational analysis, and regression analysis were performed.

II. SIGNIFICANCE OF THE STUDY

A comprehensive analysis of the impact of the COVID-19 pandemic on the happiness of people from different regions of the world was performed using the 2021 data from the World Happiness Report. This study explored the happiness score of people at the onset of the pandemic. With this study, organizations can reevaluate the effectiveness of policies and interventions that were immediately implemented to mitigate the negative impact of the pandemic and to enhance the happiness of people in their communities. Furthermore, the insights and recommendations from this study can contribute to the study of the happiness and well-being of people.

III. THE PROBLEM

The COVID-19 pandemic has significantly impacted the well-being of individuals across the globe. To understand the severity and the mechanisms of this effect, a comprehensive analysis of happiness is necessary. This study aims to fill this gap by employing various statistical techniques to analyze the 2020-2021 data from the World Happiness Report.

IV. SCOPE AND LIMITATIONS

The effect of the onset of the COVID-19 pandemic on the happiness of different regions in the world was analyzed using the 2020-2021 data of the World Happiness Report.

The report primarily sourced the data from the Gallup World Poll and incorporates it with data from other sources such as the World Risk Poll, the COVID Data Hub, and the Oxford COVID-19 Government Response Tracker for the 2021 report. The 2020 data is based on surveys and reports conducted in 2019 and early 2020 while the 2021 data is based on surveys and reports conducted in 2020 and early 2021. With the data, the following research questions were answered:

1. How does the happiness of each region vary in 2021 data compared to 2020 data and other regions?
2. What factors are associated with the change in the happiness of each region?
3. How well can the change in the happiness of each region be predicted by the selected factors?

Note that the data may not be able to capture the full impact and the long-term consequence of the pandemic as the situation may have changed since it was collected. Data quality may have also been affected by disruptions due to the pandemic. Happiness is also subjective and may not be able to be measured fully in 1 score. The happiness score of each country is obtained from a representative sample of the population so, it may contain biases due to cultural differences, social norms, environment, etc. The data may not include all factors that influence happiness. Therefore, the results of this study may not be generalizable in other contexts or periods. Lastly, the causal relationship between the COVID-19 pandemic and well-being is complex and so, the insights from the study must be interpreted with caution.

V. STATISTICAL MODEL

In general, the happiness score was the response variable used in the statistical analysis. It measures the happiness of a specific country, where higher values indicate happier individuals. Its mean per region is a regional representation of happiness and its change from 2020 to 2021 measures the effect of the pandemic on happiness. The predictors involved are the following variables from the World Happiness Report:

1. Freedom to make life choices: the extent to which people feel they have control over their own lives
2. Generosity: average contribution of people to charity
3. Healthy life expectancy: number of years a person can expect to live in good health
4. Logged GDP per capita: logarithm of the gross domestic product per person, where higher values indicate higher income and economic development
5. Perceptions of Corruption: the extent to which people think that corruption is widespread in their country
6. Social Support: the degree to which people feel they have someone to count on in times of trouble

The selection of these predictors is based on the theoretical framework of the World Happiness Report, which suggests that they are important determinants of happiness across countries and regions. They are also expected to reflect on how people cope with the pandemic and its consequences.

Inferential statistical methods used were the following:

1. One-way ANOVA test: checks if there is a significant difference in the mean happiness score among the regions
2. Paired t-test: checks if there is a significant difference in the happiness score between 2020 and 2021 for each region
3. Pearson correlation coefficient: measures the relationship between the variables
4. Multiple linear regression model: predict the happiness score for each region based on the selected factors

VI. METHODOLOGY

The 2020 World Happiness Report uses data from the surveys conducted by the Gallup World Poll in 2019. The 2021 World Happiness Report uses data from the surveys conducted by the Gallup World Poll in 2020 and early 2021. This report is expected to reflect the impact of the COVID-19 pandemic on people's happiness. The survey asks representatives from more than 150 countries to rate their happiness and various aspects of their lives from social support to freedom.

The number of samples in this study was determined by the availability of the data from the World Happiness Report, which includes data from countries grouped into 10 regions based on the regional indicator variable. Table 1 shows the number of samples or countries for each region varies. Initially, regional indicators were only available in the 2021 World Happiness Report. By merging 2020 data with 2021 data based on the country names, regional indicators were added to the 2020 data from the 2021 data where the country names matched. Assuming that the data were normally distributed and the variances were equal, One-way Analysis of Variance (ANOVA) was done to test whether there was a significant difference in the mean happiness score among the regions.

Table 1: Number of Available Countries or Samples per Region for the years 2020 and 2021

Region	**No. of Samples**	
	2020	**2021**
Sub-Saharan Africa	36	14
Western Europe	21	19
Latin America & the Caribbean	20	11
Middle East & North Africa	17	11
Central and Eastern Europe	17	16
Commonwealth of Independent States	12	7
Southeast Asia	9	5
South Asia	7	2
East Asia	6	6
North American and ANZ	4	4

The data was then aggregated for each region by the regional indicator variable, where the mean value of the happiness score was calculated for each region. Figure 1 shows the mean happiness score per region for the years 2020 and 2021. Two-sample t-test was performed to check if there is a significant difference in the happiness score between 2020 and 2021 for each region.

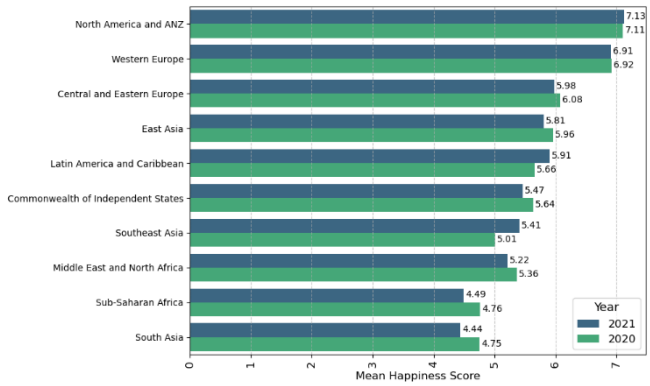

Figure 1: Mean Happiness Score per Region for the years 2020 and 2021

The correlation of each factor available in the 2021 data with the happiness score was then determined using correlation coefficients. This data set was used to fit a multiple linear regression model to predict the happiness score. Residual plots and diagnostic tests were also performed to check the assumptions and validity of the model.

VII. DATA ANALYSIS & DISCUSSIONS

A. Regional variation in mean happiness score

Table 2 shows the results of the one-way ANOVA test that was conducted to compare the mean happiness scores across the regions for 2020 and 2021. The null hypothesis was that there was no difference in the mean happiness score among the regions, and the alternative hypothesis was that there was at least one region that had a different mean happiness score from the others. The p-value for both years was less than 0.05. Therefore, the null hypothesis was rejected, and it was concluded that there was a statistically significant difference in the mean happiness score among the regions for both years. The F-statistic for both years was large, which indicates that the variation in the mean happiness score between the regions was much greater than the variation within the regions. This suggests that the regions had different happiness scores and that the region was an important factor in explaining the happiness score.

Table 2: One-Way ANOVA test result

Year	F statistic	P-value
2020	14.69	4.59e-14
2021	25.34	2.57e-25

B. Variation in happiness score across regions over time

Table 3 shows the results of the paired t-test that was performed to compare the happiness score between 2020 and 2021 for each region. The null hypothesis was that the mean difference in the happiness score between the two years within each region was equal to zero, and the alternative hypothesis was that the mean difference in the happiness score between the two years within each region was equal to zero. The p-value for all regions except Latin America and the Caribbean region was greater than 0.05, which indicates that the observed mean difference in the happiness score between the two years within each region was not statistically significant at the 5% level of significance. Therefore, the null hypothesis could not be rejected, and it was concluded that there was insufficient evidence to claim that there was a statistically significant difference in the happiness score between the two years within each region. The absolute value of the t-statistic for all regions except Latin America and the Caribbean region was less than two, which means that the observed mean difference was within two standard errors of zero. This suggests that the observed mean difference was not very large and that the year was not an important factor in explaining the change in the happiness score.

Table 3: Two sample t-test results

Region	T statistic	P-value
Sub-Saharan Africa	1.06	0.31
Western Europe	-1.52	0.15
Latin America & the Caribbean	-5.00	5.37e-4
Middle East & North Africa	-0.87	0.41
Central and Eastern Europe	1.90	7.67e-2
Commonwealth of Independent States	1.90	0.11
Southeast Asia	-1.19	0.30
South Asia	4.38	0.14
East Asia	1.55	0.18
North American and ANZ	-0.50	0.65

Figure 2 shows the mean happiness score difference between 2020 and 2021 for each region. The mean happiness score difference was calculated by subtracting the happiness score in 2020 from the happiness score in 2021 for each country and then getting the mean of the difference for each region. A positive mean happiness score difference indicates that the region had a higher happiness score in 2021 than in 2020, while a negative mean happiness score difference indicates the opposite. The only region that had a statistically significant mean happiness score difference, as indicated by Table 3, was Latin America and the Caribbean region. The mean happiness score difference for this region was 0.26, which means that the region had a slightly higher happiness score in 2021 than in 2020, contrary to what was expected for it to be due to the pandemic.

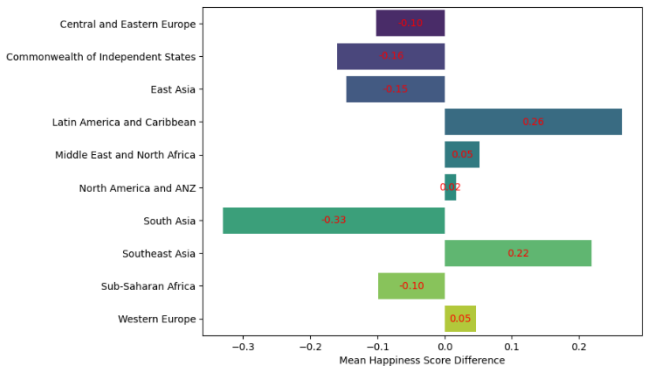

Figure 2: Mean Happiness Score Difference per Region for the years 2020 and 2021

C. *Correlation between other factors and the happiness score at the onset of the pandemic*

Table 4 shows the Pearson correlation coefficient to measure the linear relationship between the factors and the happiness score. It indicates that the logarithm of gross domestic product per capita, social support, and healthy life expectancy has a strong positive correlation with the happiness score, respectively; freedom to make life choices has a moderate positive correlation; perception of corruption has a moderate negative correlation; and generosity does not correlate. This suggests that the logarithm of gross domestic product per capita, social support, and healthy life expectancy are influential while generosity, which does not correlate, can be removed from the model.

Table 4: Pearson correlation coefficient results

Factors	**Correlation Coefficient**
Freedom to make life choices	0.54
Generosity	-0.061
Healthy life expectancy	0.74
Log GDP per capita	0.84
Perceptions of corruption	-0.51
Social support	0.77

D. *Multiple Linear Regression Model for Happiness*

Figure 3 shows the results of an ordinary least squares regression analysis of happiness score as a function of five predictors: freedom to make life choices, healthy life expectancy, logged GDP per capita, perceptions of corruption, and social support. The model excludes generosity as a predictor. It highly fits the data, with an R-squared value of 0.754 and an adjusted R-squared value of 0.745. The F-statistic of 87.49 and the associated p-value of 9.68e-42 show that the model is statistically significant and that the predictors have a joint effect on happiness score.

The model can be expressed by the following equation:

$$y = 2.132\,x_1 + 0.029\,x_2 + 0.264\,x_3 - 0.668\,x_4 + 2.507\,x_5 - 2.110$$

where y is the happiness score, x_1 is the freedom to make life choices, x_2 is the healthy life expectancy, x_3 is the logged GDP per capita, x_4 is the perceptions of corruption, and x_5 is the social support. All predictors have positive and significant coefficients, except for the perception of corruption, which has a negative and significant coefficient. This implies that higher values of freedom, health, GDP, and social support are associated with higher happiness scores, while higher values of corruption are associated with lower happiness scores. The standard errors of the coefficients are smaller than the absolute values of the coefficients, indicating that the predictors have a non-zero effect on happiness scores. The t-values and the p-values of the coefficients confirm that the predictors are individually significant at the 0.05 level.

The model, however, may not be the best fit for the data, as the residuals are not normally distributed. The omnibus test statistic of 13.817 and the corresponding p-value of 0.001 indicate that the null hypothesis of normality is rejected. The skewness of -0.687 and the kurtosis of 3.727 suggest that the distribution of the residuals is left-skewed and has a longer left tail and a higher peak than a normal distribution. The Jarque-Bera test statistic of 14.955 and the associated p-value of 5.66e-4 also reject the null hypothesis of normality. Therefore, the model may not meet the assumption of normality of errors, which could affect the validity of the inference.

OLS Regression Results

Dep. Variable:	Ladder score	R-squared:	0.754
Model:	OLS	Adj. R-squared:	0.745
Method:	Least Squares	F-statistic:	87.49
Date:	Fri, 26 Jan 2024	Prob (F-statistic):	9.68e-42
Time:	07:24:41	Log-Likelihood:	-117.18
No. Observations:	149	AIC:	246.4
Df Residuals:	143	BIC:	264.4
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-2.1104	0.621	-3.398	0.001	-3.338	-0.883
Freedom to make life choices	2.1327	0.483	4.412	0.000	1.177	3.088
Healthy life expectancy	0.0294	0.013	2.204	0.029	0.003	0.056
Logged GDP per capita	0.2640	0.086	3.075	0.003	0.094	0.434
Perceptions of corruption	-0.6678	0.285	-2.339	0.021	-1.232	-0.103
Social support	2.5067	0.668	3.751	0.000	1.186	3.828

Omnibus:	13.817	Durbin-Watson:	1.590
Prob(Omnibus):	0.001	Jarque-Bera (JB):	14.955
Skew:	-0.687	Prob(JB):	0.000566
Kurtosis:	3.723	Cond. No.	1.14e+03

Figure 3: Ordinary Least Squares Regression Result for Happiness

VIII. CONCLUSION

This study analyzed the effect of the COVID-19 pandemic on the happiness of different regions of the world, using the data from the World Happiness Report for 2020 and 2021. We found that the happiness score varied significantly across regions, but not over time from 2020 to 2021, except for Latin America and the Caribbean region, which showed a slight increase in mean happiness score in 2021. We also found that the happiness score was strongly influenced by logged GDP per capita, social support, and healthy life expectancy, moderately affected by the freedom to make life choices and perceptions of corruption, and not affected by generosity. We also used a multiple linear regression model to quantify the relationship between the happiness score and the factors, but the model did not satisfy the normality assumption of errors, which could limit the reliability of the results. Therefore, the study suggested that a better model that meets the normality assumption of errors should be developed and applied to the data. The study contributed to the literature on happiness and well-being by providing a regional analysis of the effect of the COVID-19 pandemic on the happiness of people.

APPENDIX

Raw data can be downloaded at: [World Happiness Report 2021 (kaggle.com)]

Python Script and raw data used for statistical analysis can be checked at: [MNDC12/IE211_Happiness_Report: Mini Project for IE (github.com)]

REFERENCES

[1] World Health Organization 2023 data.who.int, WHO Coronavirus (COVID-19) dashboard > Deaths [Dashboard]. https://data.who.int/dashboards/covid19/deaths

[2] World Bank. World Development Report 2022: Finance for an equitable recovery. The World Bank, 2022.

[3] Sibley, C. G., et al. "Short-term effects of the COVID-19 pandemic and a nationwide lockdown on institutional trust, attitudes to government, health and wellbeing." American Psychological Association 75.5 (2020): 618-663.

[4] Brodeur, Abel, et al. "Assessing the impact of the coronavirus lockdown on unhappiness, loneliness, and boredom using Google Trends." arXiv preprint arXiv:2004.12129 (2020).

[5] Helliwell, John F., et al. "World happiness report 2021." (2021).

[6] Helliwell, John F., et al. "World happiness report 2020." (2020).