Assignment 1: World Happiness Score Prediction

By Gyana Deepika Dasara [gd452]

July 22, 2023

1 Introduction

The pursuit of happiness and well-being has gained significant recognition worldwide as essential components of societal progress and development. On 28 June 2012, the United Nations General Assembly adopted Resolution 66/281, designating 20 March as the International Day of Happiness. Aligned with this celebration, the World Happiness Report is released annually around March 20th.

The happiness score/index holds significant value and offers numerous benefits. Predicting it is crucial for policymakers, enabling them to make informed decisions that enhance citizen well-being. It facilitates efficient resource allocation by targeting interventions to areas with lower predicted happiness scores, ensuring effective utilization. Moreover, it enables comparative analysis across countries, providing valuable insights into the factors that shape happiness variations and guiding the adoption of successful strategies on a global scale.

2 Project Proposal

In this project, my aim is to delve into the factors that contribute to the calculation of the happiness score, a measure of overall well-being. I will focus on variables such as GDP per capita, healthy life expectancy, population, social support, freedom to make life choices, Covid-19 deaths, and perceptions of corruption.

Through the use of multi-linear regression, I will predict the happiness score based on these variables. Additionally, I will employ correlation techniques to explore any relationships that may exist among these variables. The unit of analysis for this project will be country-wise, providing insights into the happiness factors on a national level.

2.1 Data Sources

For this project, I will be utilizing data from two distinct sources to gather comprehensive information.

The first source is the worldhappiness.report website, which provides valuable data on the happiness index, as well as other significant variables such as GDP per capita, social support, healthy life expectancy, freedom, generosity, and corruption. The dataset from the World Happiness Report encompasses data from 137 countries, making it a robust and diverse source of information.

The second source I will be utilizing is the website called worldometers. To obtain population count and Covid deaths data for various countries, I will employ web scraping techniques to gather the required information. This data will be stored in a convenient CSV file format, enabling easy manipulation and analysis. By merging the data from both sources, I will create a comprehensive dataset that combines happiness-related variables with country population information.

By combining the datasets from these two sources, we can gain a richer understanding of the factors influencing happiness while also considering the population dynamics of different countries.

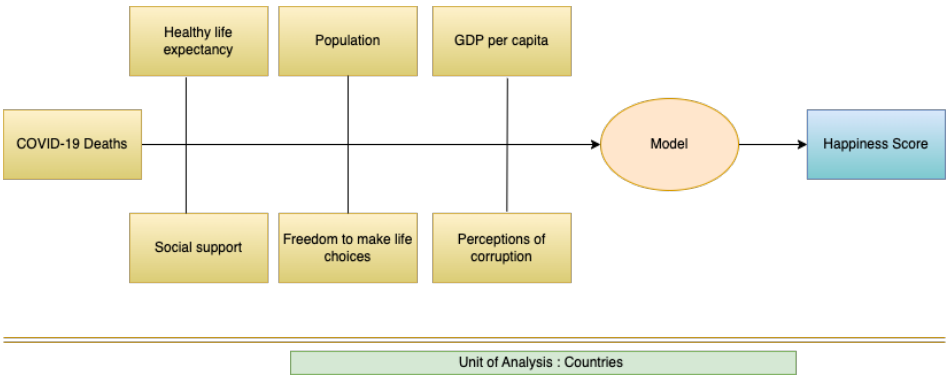

Figure 1: Model Overview

2.2 Model

We will construct a multi-linear regression model to explore the relationship between several independent variables and the 'Happiness Score' as the outcome. The model will utilize a set of carefully selected independent variables (x) to predict the happiness score accurately.

The independent variables (x) considered in our analysis are as follows:

1. **GDP per capita**: This variable reflects the economic well-being of a country, measuring the average income per person.
2. **Healthy life expectancy**: It represents the average number of years individuals can expect to live in good health, capturing the well-being of the population.
3. **Population**: This variable accounts for the size of the population in a country, providing insights into the scale and demographic context.
4. **Social support**: It signifies the presence of social networks and supportive relationships within a society, indicating the level of social well-being.
5. **Freedom to make life choices**: This variable assesses the extent to which individuals have the freedom and autonomy to make decisions that impact their lives.
6. **Perceptions of corruption**: It measures the level of perceived corruption within a country, influencing the overall trust and well-being of its citizens.
7. **Covid-19 Deaths**: It is the total number of deaths due to Covid-19 per population in a country.

The dependent variable (y) is:

- **Happiness Score**

By considering these independent variables in our multi-linear regression model, we aim to unravel their collective influence on the happiness score, providing valuable insights into the factors that contribute to overall happiness.

The regression equation for our analysis will be formulated as follows:

$$\begin{aligned}\text{HappinessScore}_i = \alpha &+ \beta_1(\text{GDP.per.capita})_i \\ &+ \beta_2(\text{Healthy.life.expectancy})_i \\ &+ \beta_3(\text{Population})_i \\ &+ \beta_4(\text{Social.support})_i \\ &+ \beta_5(\text{Freedom})_i \\ &+ \beta_5(\text{Covid19.Deaths})_i \\ &+ \beta_6(\text{Perceptions.of.corruption})_i + \epsilon\end{aligned}$$

2.3 Informed Estimate of Findings

The chosen independent variables of GDP, health, population, social support, freedom, and corruption have been selected to predict the happiness score due to their significant impact on individual well-being. Higher GDP should reflect better economic conditions and material well-being, providing access to resources and improved living standards.

Good health, including access to quality healthcare and longer life expectancy, contributes to overall happiness and fulfillment. Population must influence the happiness index through social interactions, diversity, economic opportunities, and access to services and infrastructure. There is a potential inverse relationship between COVID deaths and the happiness index of a country. This is because the loss of family members, friends, and loved ones due to COVID-19 can have a profound emotional and psychological impact on individuals, which can adversely affect their overall happiness and well-being.

Social support, derived from relationships with family, friends, and communities, plays a crucial role in fostering a sense of belonging and well-being. Freedom in various aspects of life, such as personal choices, expression, and opportunities, empowers individuals to pursue their goals and experience higher satisfaction and happiness.

Lastly, the presence of lower levels of corruption, which undermines trust and fairness, should be associated with higher happiness levels as citizens feel secure and have confidence in their institutions.

By considering these variables, I aim to capture the multidimensional factors that influence happiness, enabling a comprehensive understanding of the predictors and their impact on individuals and societies.

2.4 Challenges

Predicting the happiness index can present several challenges:

1. **Subjectivity**: Happiness is a subjective experience, and measuring it accurately can be difficult. People's perceptions and definitions of happiness can vary, making it challenging to develop standardized metrics and indicators.
   To address this, I am relying on previous research to inform the selection of variables that are strongly associated with happiness. By drawing on established findings, I aim to reduce subjectivity and ensure a more objective approach in determining the factors that contribute to overall well-being.

2. **Multidimensional nature**: Happiness is influenced by various factors, including economic, social, psychological, and cultural aspects. Incorporating and weighting these factors appropriately in predictive models can be complex due to their interrelated and multifaceted nature.

3. **Data availability and quality**: Obtaining reliable and comprehensive data on all relevant variables can be challenging. Data collection methods, sample sizes, and data quality can vary across countries, which may affect the accuracy and generalizability of predictions.

4. **External factors and dynamics**: Happiness can be influenced by external factors such as political events, economic crises, or natural disasters. Incorporating these dynamic factors and their impacts on happiness predictions requires continuous data updates and sophisticated modeling techniques. To address this, I am introducing a new variable, namely "COVID-19 deaths," to explore whether a recent pandemic has influenced the happiness levels of a nation.

2.5 Data Curation

The Excel Book is structured into four sheets. The first sheet holds the final curated data, which will serve as the foundation for my subsequent modeling and analysis. In this sheet, the yellow row represents the countries, which serve as the unit of analysis. Furthermore, it is the primary column I've utilized to merge features from other datasets. The blue columns contain information extracted from the world happiness report (data source 1), while the pink columns represent data collected from the worldometer website (data source 2). The green column denotes the dependent variable of interest that I aim to predict.

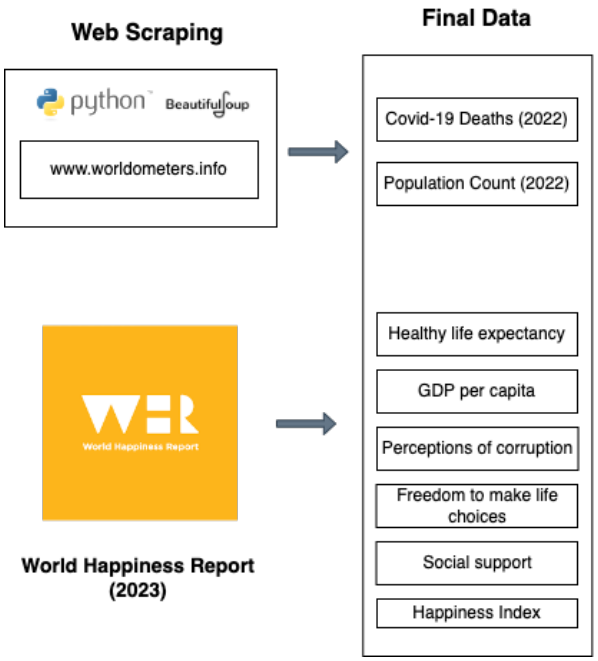

Figure 2: Data Curation

Moving on to the second sheet, it contains the original datasets sourced from the worldhappiness.report/ website. The remaining two sheets encompass tables acquired through web scraping techniques (Python & BeautifulSoup Library) from the worldometer website. These tables specifically pertain to population and COVID-19 death statistics.

In order to enhance the analysis, I have taken a more comprehensive approach by calculating the percentage of COVID deaths per population for the given year, rather than solely relying on the count of COVID-19 deaths as a feature. This metric provides a fair and standardized measure of the virus's impact across diverse populations, enabling more meaningful comparisons and informed decision-making.

Furthermore, I have addressed missing values and inconsistencies in the data related to country names. For missing values, I have sought alternative sources to populate the population and COVID-19 death information where necessary. In cases where a row pertained to a specific province within a country and lacked relevant information, I made the decision to exclude that particular row. To tackle inconsistencies in country names across the three datasets, I have standardized the country names and manually rectified any discrepancies, ensuring consistency across all datasets. This step is crucial, as the country name serves as a key metric for merging features from different datasets to obtain the final curated data.

The final dataset consists of 134 rows and 9 columns, with the country name being the only non-numeric column. All other columns are numeric and will be utilized to predict the happiness index feature.

To clarify, I have accessed the 2023 world happiness report and retrieved population and COVID-19 death data for the year 2022. The 2022 data is used in calculating the happiness scores for the 2023 report.

Raw Data:

1. Worldhappiness.report : Country name, Ladder score, Standard error of ladder score, upperwhisker, lowerwhisker, Logged GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity, Perceptions of corruption, Ladder score in Dystopia, Explained

Country name	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Perceptions of corruption	Population (2022)	Covid-19 Deaths per population (2022)	Happiness Index
Finland	10.792	0.969	71.15	0.961	0.182	5540720	0.1800127059	7.804
Denmark	10.962	0.954	71.25	0.934	0.196	5792202	0.1511514964	7.586
Iceland	10.896	0.983	72.05	0.936	0.668	341243	0.06710760367	7.53
Israel	10.639	0.943	72.697	0.809	0.708	8655535	0.1451556721	7.473
Netherlands	10.942	0.93	71.55	0.887	0.379	17134872	0.1341825022	7.403
Sweden	10.883	0.939	72.15	0.948	0.202	10099265	0.242948373	7.395
Norway	11.088	0.943	71.5	0.947	0.283	5421241	0.1026702189	7.315
Switzerland	11.164	0.92	72.9	0.891	0.266	8654622	0.166985918	7.24
Luxembourg	11.66	0.879	71.675	0.915	0.345	625978	0.1968120285	7.228
New Zealand	10.662	0.952	70.35	0.887	0.271	4822233	0.09400209405	7.123
Austria	10.899	0.888	71.15	0.855	0.497	9006398	0.2502887392	7.097
Australia	10.821	0.934	71.05	0.89	0.496	25499884	0.08555725195	7.095
Canada	10.773	0.929	71.4	0.874	0.42	37742154	0.1400555994	6.961
Ireland	11.527	0.905	71.3	0.874	0.358	4937786	0.1834425388	6.911
United States	11.048	0.919	65.85	0.8	0.689	331002651	0.3529512517	6.894
Germany	10.879	0.896	71.3	0.846	0.42	83783942	0.2080971554	6.892
Belgium	10.844	0.915	70.899	0.825	0.549	11589623	0.2966015374	6.859
Czechia	10.611	0.953	69.05	0.903	0.859	10708981	0.3997672608	6.845
United Kingdom	10.704	0.882	70.3	0.852	0.454	67886011	0.3351559425	6.796
Lithuania	10.568	0.939	67.397	0.748	0.805	2722289	0.3560239196	6.763
France	10.701	0.909	72.3	0.819	0.553	65273511	0.2568300639	6.661
Slovenia	10.588	0.951	71.052	0.913	0.771	2078938	0.3415205263	6.65
Costa Rica	9.952	0.872	70	0.895	0.768	5094118	0.1842517193	6.609
Romania	10.339	0.848	67.051	0.856	0.929	19237691	0.3547203248	6.589
Singapore	11.571	0.878	73.8	0.878	0.146	5850342	0.03146824579	6.587

Figure 3: Final Curated Data

by: Log GDP per capita, Explained by: Social support, Explained by: Healthy life expectancy, Explained by: Freedom to make life choices, Explained by: Generosity, Explained by: Perceptions of corruption, Dystopia + residual.

2. Population Data (worldometer) : Country name, Population (2022), Yearly Change, Net Change, Density (P/Km^2), Land Area (km^2), Migrants (net), Fert. Rate, Med. Age, Urban Pop., % World Share.

3. Covid-19 Data (worldometer) : Country name, Total Cases New Cases, Total Deaths, New Deaths, Total Recovered, New Recovered, Active Cases, Serious, Critical, Tot Cases/ 1M pop, Deaths/ 1M pop, Total Tests, Tests/ 1M pop, Population, Continent, 1 Case every X ppl, 1 Death every X ppl, 1 Test every X ppl, New Cases/1M pop, New Deaths/1M pop, Active Cases/1M pop.

Final Curated Data:

- Country name, Logged GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Perceptions of corruption, Population (2022), Covid-19 Deaths per population (2022), Happiness Index.

3 SPSS Analysis

3.1 Correlation

3.1.1 Correlation Scores

- GDP per capita & Happiness Score: r = 0.779
- Social support & Happiness Score: r = 0.839
- Healthy life expectancy & Happiness Score: r = 0.744
- Freedom to make life choices & Happiness Score: r = 0.658
- Perceptions of corruption & Happiness Score: r = -0.477
- Population (2022) & Happiness Score: r = -0.80
- Covid-19 Deaths per population (2022) & Happiness Score: r = 0.408

Correlations

		HappinessIndex	LoggedGDPpercapita	Socialsupport	Healthylifeexpectancy	Freedomtomakelifechoices	Perceptionsofcorruption	Population2022	Covid19Deathsperpopulation2022
HappinessIndex	Pearson Correlation	1	.779**	.839**	.744**	.658**	-.477**	-.080	.408**
	Sig. (2-tailed)		<.001	<.001	<.001	<.001	<.001	.360	<.001
	N	134	134	134	133	134	134	134	134
LoggedGDPpercapita	Pearson Correlation	.779**	1	.737**	.835**	.443**	-.436**	-.027	.518**
	Sig. (2-tailed)	<.001		<.001	<.001	<.001	<.001	.757	<.001
	N	134	134	134	133	134	134	134	134
Socialsupport	Pearson Correlation	.839**	.737**	1	.717**	.537**	-.273**	-.087	.508**
	Sig. (2-tailed)	<.001	<.001		<.001	<.001	.001	.319	<.001
	N	134	134	134	133	134	134	134	134
Healthylifeexpectancy	Pearson Correlation	.744**	.835**	.717**	1	.406**	-.405**	-.033	.513**
	Sig. (2-tailed)	<.001	<.001	<.001		<.001	<.001	.708	<.001
	N	133	133	133	133	133	133	133	133
Freedomtomakelifechoices	Pearson Correlation	.658**	.443**	.537**	.406**	1	-.386**	.097	.125
	Sig. (2-tailed)	<.001	<.001	<.001	<.001		<.001	.265	.151
	N	134	134	134	133	134	134	134	134
Perceptionsofcorruption	Pearson Correlation	-.477**	-.436**	-.273**	-.405**	-.386**	1	.041	.109
	Sig. (2-tailed)	<.001	<.001	.001	<.001	<.001		.635	.208
	N	134	134	134	133	134	134	134	134
Population2022	Pearson Correlation	-.080	-.027	-.087	-.033	.097	.041	1	-.119
	Sig. (2-tailed)	.360	.757	.319	.708	.265	.635		.171
	N	134	134	134	133	134	134	134	134
Covid19Deathsperpopulation2022	Pearson Correlation	.408**	.518**	.508**	.513**	.125	.109	-.119	1
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	.151	.208	.171	
	N	134	134	134	133	134	134	134	134

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 4: Correlations between dependent and independent variables

3.1.2 Significance Values

- GDP per capita & Happiness Score: p <0.01 (Statistically Significant)
- Social support & Happiness Score: p <0.01 (Statistically Significant)
- Healthy life expectancy & Happiness Score: p <0.01 (Statistically Significant)
- Freedom to make life choices & Happiness Score: p <0.01 (Statistically Significant)
- Perceptions of corruption & Happiness Score: p <0.01 (Statistically Significant)
- Population (2022) & Happiness Score: p = 0.360, which is not <0.05 (Not Statistically Significant)
- Covid-19 Deaths per population (2022) & Happiness Score: p <0.01 (Statistically Significant)

3.1.3 Correlation Directions

- GDP per capita & Happiness Score: Positive Direction
- Social support & Happiness Score: Positive Direction
- Healthy life expectancy & Happiness Score: Positive Direction
- Freedom to make life choices & Happiness Score: Positive Direction
- Perceptions of corruption & Happiness Score: Negative Direction
- Population (2022) & Happiness Score: Negative Direction
- Covid-19 Deaths per population (2022) & Happiness Score: Positive Direction

3.1.4 Correlation: Interpretation & Analysis

- The correlation (r) value of GDP per capita and Happiness Score produced was positive and close to one. This would indicate a positive direction, hinting that the dependent variable (Happiness Score) increases when the independent variable (in this instance, GDP per capita) increases. The higher absolute r-value indicates a more linear relationship. The p-value derived from the relationship between these features was less than the 0.05 critical value. This implies that there is a statistical significance and a greater likelihood of this value being a genuine correlation, rather than a coincidence.

- The correlation (r) value of Social support and Happiness Score produced was positive and close to one. This would indicate a positive direction, hinting that the dependent variable (Happiness Score) increases when the independent variable (in this instance, Social support) increases. The higher absolute r-value indicates a more linear relationship. The p-value derived from the relationship between these features was less than the 0.05 critical value. This implies that there is a statistical significance and a greater likelihood of this value being a genuine correlation, rather than a coincidence.

- The correlation (r) value of Healthy life expectancy and Happiness Score produced was positive and close to one. This would indicate a positive direction, hinting that the dependent variable (Happiness Score) increases when the independent variable (in this instance, Healthy life expectancy) increases. The higher absolute r-value indicates a more linear relationship. The p-value derived from the relationship between these features was less than the 0.05 critical value. This implies that there is a statistical significance and a greater likelihood of this value being a genuine correlation, rather than a coincidence.

- The correlation (r) value of Freedom to make life choices and the Happiness Score produced was positive. This would indicate a positive direction, hinting that the dependent variable (Happiness Score) increases when the independent variable (in this instance, Freedom to make life choices) increases. The higher absolute r-value indicates a more linear relationship. The p-value derived from the relationship between these features was less than the 0.05 critical value. This implies that there is a statistical significance and a greater likelihood of this value being a genuine correlation, rather than a coincidence.

- The correlation (r) value of Perceptions of corruption and the Happiness Score produced was negative and closer to zero. This would indicate a negative direction, hinting that the dependent variable (Happiness Score) decreases when the independent variable (in this instance, Perceptions of corruption) increases. The lower r-value indicates a weaker linear relationship that is more scattered. The p-value derived from the relationship between these features was less than the 0.05 critical value. This implies that there is a statistical significance and a greater likelihood of this value being a genuine correlation, rather than a coincidence.

- The correlation (r) value of Population and the Happiness Score produced was negative. This would indicate a negative direction, hinting that the dependent variable (Happiness Score) decreases when the independent variable (in this instance, Population) increases. The p-value derived from the relationship between these features was more than the 0.05 critical value. This implies that there is no statistical significance and a greater likelihood of this value being a coincidence, rather than a genuine correlation.

- The correlation (r) value of Covid-19 Deaths per population and the Happiness Score produced was positive and closer to zero. This would indicate a positive direction, hinting that the dependent variable (Happiness Score) increases when the independent variable (in this instance, Covid-19 Deaths per population) increases. The p-value derived from the relationship between these features was less than the 0.05 critical value. This implies that there is a statistical significance and a greater likelihood of this value being a genuine correlation, rather than a coincidence. This is surprising.

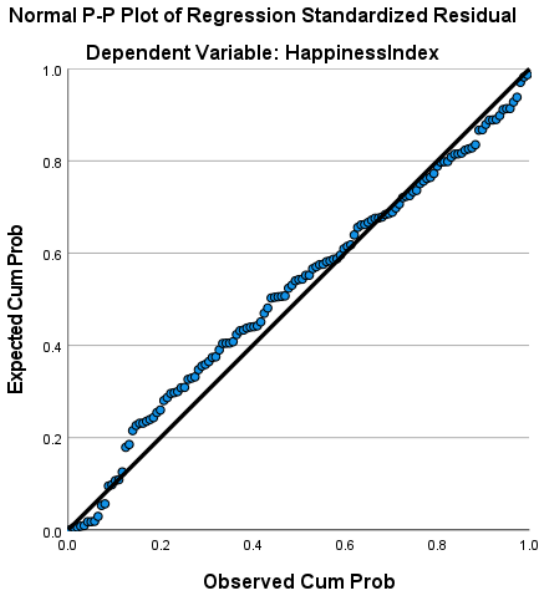

Figure 5: Model Plot

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.913[a]	.834	.825	.475272

a. Predictors: (Constant), Covid19Deathsperpopulation2022, Perceptionsofcorruption, Population2022, Freedomtomakelifechoices, Healthylifeexpectancy, Socialsupport, LoggedGDPpercapita

b. Dependent Variable: HappinessIndex

Figure 6: Multiple Regression Model Summary

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-2.012	.802		-2.509	.013
	LoggedGDPpercapita	.172	.071	.183	2.432	.016
	Socialsupport	4.027	.556	.457	7.248	<.001
	Healthylifeexpectancy	.022	.014	.112	1.594	.113
	Freedomtomakelifechoices	2.378	.467	.237	5.095	<.001
	Perceptionsofcorruption	-.836	.300	-.132	-2.786	.006
	Population2022	-3.200E-10	.000	-.050	-1.323	.188
	Covid19Deathsperpopulation2022	.010	.389	.001	.026	.979

a. Dependent Variable: HappinessIndex

Figure 7: Model Coefficients

3.2 Multiple Linear Regression

3.2.1 Percentage of Variance explained by the created model

R-Square = 0.834 = 83.4%

- 83.4% of the variance is explained by the model.

3.2.2 Significant Variables

- GDP per capita & Happiness Score: p=0.016 <0.05 (Statistically Significant)
- Social support & Happiness Score: p <0.05 (Statistically Significant)
- Healthy life expectancy & Happiness Score: p=0.113 <0.05 (Not Statistically Significant)
- Freedom to make life choices & Happiness Score: p <0.05 (Statistically Significant)
- Perceptions of corruption & Happiness Score: p=0.006 <0.05 (Statistically Significant)
- Population (2022) & Happiness Score: p = 0.188, which is not <0.05 (Not Statistically Significant)
- Covid-19 Deaths per population (2022) & Happiness Score: p=0.979 <0.05 (Not Statistically Significant)

3.2.3 Coefficients

- **Unstandardized Coefficients**
  An increase of the independent variables by an index of one will reflect a change in index for the dependent (outcome) variable in the amount of the specified unstandardized coefficients.
  - **GDP per capita**: 0.172 (Positive) - For every 1 unit of change in GDP per capita, the dependent variable (Happiness Index), increases by 0.172. 5
  - **Social support**: 4.027 (Positive) - For every 1 unit of change in Social support, the dependent variable (Happiness Index), increases by 4.027.
  - **Healthy life expectancy**: 0.22 (Positive) - For every 1 unit of change in Healthy life expectancy, the dependent variable (Happiness Index), increases by 0.22.
  - **Freedom to make life choices**: 2.378 (Positive) - For every 1 unit of change in Freedom to make life choices, the dependent variable (Happiness Index), increases by 2.378.

- **Perceptions of corruption**: -0.836 (Negative) - For every 1 unit of change in Perceptions of corruption, the dependent variable (Happiness Index), decreases by -0.836.
- **Population**: -3.200E-10 (Negative) - For every 1 unit of change in Population, the dependent variable (Happiness Index), decreases by -3.200E-10.
- **Covid-19 Deaths per population**: 0.010 (Positive) - For every 1 unit of change in Covid-19 Deaths per population, the dependent variable (Happiness Index), increases by 0.010.

- **Standardized Coefficients**
  An increase of the independent variables by a standard deviation of one will reflect a change in standard deviation for the dependent (outcome) variable in the amount of the specified standardized coefficients.
  - **GDP per capita**: 0.183 (Positive) - For every 1 standard deviation of movement in GDP per capita, the dependent variable (Happiness Index), increases by 0.183.
  - **Social support**: 0.457 (Positive) - For every 1 standard deviation of movement in Social support, the dependent variable (Happiness Index), increases by 0.457.
  - **Healthy life expectancy**: 0.112 (Positive) - For every 1 standard deviation of movement in Healthy life expectancy, the dependent variable (Happiness Index), increases by 0.112.
  - **Freedom to make life choices**: 0.237 (Positive) - For every 1 standard deviation of movement in Freedom to make life choices, the dependent variable (Happiness Index), increases by 0.237.
  - **Perceptions of corruption**: -0.132 (Negative) - For every 1 standard deviation of movement in Perceptions of corruption, the dependent variable (Happiness Index), decreases by -0.132.
  - **Population**: -0.50 (Negative) - For every 1 standard deviation of movement in Population, the dependent variable (Happiness Index), decreases by -0.50.
  - **Covid-19 Deaths per population**: 0.001 (Positive) - For every 1 standard deviation of movement in Covid-19 Deaths per population, the dependent variable (Happiness Index), increases by 0.001.

3.2.4 Interpretations, Analysis, & Policy Making

The model summary reveals a promising R-square value of 83.4%, indicating that a significant portion of the variance in the Happiness index can be explained by the independent variables. Upon examining the regression line fit, we find that it aligns well with the data.

In the bivariate correlation results, all independent variables, except for the population feature, demonstrated statistical significance. Notably, Logged GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Perceptions of corruption, and Covid-19 Deaths per population (2022) exhibited significant correlations.

However, in the model correlation results, Healthy life expectancy, Population (2022), and Covid-19 Deaths per population did not show statistical significance. The p-value for Healthy life expectancy, at 0.113, exceeds 0.05, suggesting a possibility of coincidence rather than a correlation. To further explore its significance, we could consider utilizing a larger dataset or employing techniques to fine-tune the model. In my opinion, Healthy life expectancy should be significant, as physical health positively influences mental well-being, contributing to overall happiness. On the other hand, the p-value for Covid-19 Deaths per population is 0.979, well above 0.05, indicating its statistical insignificance. Combining data from both pre and post-Covid times could shed light on the true effect of this feature on the happiness index. It's important to note that the insignificance of the Population in the previous bivariate analysis persists. Given the limited dataset of just 134 rows, drawing definitive conclusions may not be appropriate.

Examining the coefficients, it appears that for both standardized and unstandardized coefficients, the Perceptions of corruption and Population independent variables possessed a negative relationship. This implies that the dependent Happiness Index variable decreases in index (for the unstandardized coefficient) and standard deviation (for the standardized coefficient) for each independent variable increase in these relationships. These negative relationships suggest that higher levels of perceived

corruption and larger population size are associated with lower levels of happiness on average, as observed in the dataset used for the regression model.

For the other independent variables (Logged GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Covid-19 Deaths per population (2022)), both the standardized and unstandardized coefficients are positive. The positive coefficients for the Covid-19 Deaths per population (2022) variable in both the standardized and unstandardized coefficients are surprising and may require further investigation. The expectation would be that Covid-19 Deaths per population should have an inverse relationship with the dependent variable (Happiness Index), as higher Covid-19 deaths might be associated with lower happiness levels due to the impact of the pandemic on people's well-being and overall quality of life. Several reasons could account for this phenomenon. Firstly, the data used for the analysis might encompass a specific period where the Covid-19 pandemic had a distinct impact on happiness. Factors such as the stage of the pandemic, vaccination rates, and government responses could influence this relationship. Additionally, there might be lag effects between Covid-19 deaths and changes in happiness levels. The immediate impact of deaths may not be fully reflected in the same year, leading to discrepancies in the results.

As policy makers delve into the results, several vital considerations emerge to shape effective policies and decision-making. These aspects include:

- Equity and Inclusion: Delve into how these variables impact diverse segments of the population. Prioritize policies that reduce well-being disparities and foster happiness across various demographic groups.
- Balancing Short-Term and Long-Term Goals: Ponder the time frame for policy interventions' impact. Recognize that some variables yield short-term effects (e.g., GDP per capita), while others necessitate sustained investments for long-term gains (e.g., healthy life expectancy).
- Interdisciplinary Insights: Explore the interconnectedness of policies across different domains. Seek synergistic policies that collectively bolster overall well-being and happiness.
- Cultural and Contextual Relevance: Embrace policies tailored to each country's unique cultural, social, and economic fabric. Appreciate that what works well in one nation may not align with differing values and norms elsewhere.

Through attentive consideration of these facets, policy makers can fashion a comprehensive and responsive approach to promoting happiness and improving the lives of their constituents.

3.3 Ethical Implications, Study Limitations, & Future Research

The foundation of this analysis rests upon data sourced from the esteemed World Happiness Report and the Worldometer website. The World Happiness Report draws upon data collected through the Gallup World Poll surveys. With each passing year, the number of people and countries surveyed varies, culminating in the participation of over 100,000 individuals from 130 countries. A typical annual sample encompasses 1,000 respondents from each country.

3.3.1 Study Limitations

As with any dataset, there are essential facets that warrant our thoughtful attention. Addressing these aspects is crucial, encompassing both data collection intricacies and ethical considerations.

Here are some limitations in our data and analysis:

- Sampling Bias: The dataset's accuracy can be affected by sampling bias, as surveys may not fully represent the diversity of the global population. The sample size may not be sufficient to capture the entire population's complexity and variations. We are using data associated with just one year which may not be enough to make any strong statements.
- Self-Reported Data: Happiness is a subjective measure, and relying on self-reported data can introduce biases due to respondents' perceptions and interpretation of their own happiness levels.

- Sample Size Variability: The variability in the number of people and countries surveyed each year can impact the dataset's accuracy and the ability to make meaningful comparisons between years.
- Representativeness: The dataset's representativeness can be affected by the choice of countries surveyed and the proportion of participants from each country. Countries with larger populations may have more significant impacts on the overall results.
- Data Collection Consistency: If data collection practices or methodologies change over time, it could introduce inconsistencies in the dataset, affecting the accuracy of trend analysis.
- Missing Features: Notably, certain important variables, such as unemployment or inequality, may be absent due to the unavailability of comparable international data for the full sample of countries.

3.3.2 Ethical Implications

The data also raises ethical considerations of this nature:

- Informed Consent: It's essential to ensure that participants in the Gallup World Poll and other surveys conducted by individual countries provide informed consent to participate voluntarily. Without proper consent, using their data could raise ethical concerns about privacy and autonomy.
- Data Privacy: Since the dataset contains personal information about respondents' well-being and happiness, it's crucial to handle the data with the utmost privacy and security to protect the individuals' identities and sensitive information.
- Data Transparency: The dataset should be transparently sourced, and data collection methodologies should be well-documented. Lack of transparency can lead to doubts about the credibility of the data and potential ethical implications.
- Cultural Sensitivity: Happiness is a subjective and culturally influenced concept. Cross-cultural variations in understanding and expressing happiness should be considered, and the dataset's analysis should be sensitive to cultural differences.

3.3.3 Future Work

Future research on predicting the happiness index of a country holds exciting potential for further advancements and deeper insights:

1. Enriching the Analysis: By incorporating crucial features such as education, employment, and social class information, we gain deeper insights into the factors that truly contribute to the happiness index.
2. Embracing Long-Term Perspectives: Shifting our focus beyond a single year's data from the World Happiness Report to encompass multiple years will undoubtedly refine our results and enrich our inferences.
3. Considering External Influences: Exploring ways to integrate external events like pandemics or natural disasters into the dataset offers a nuanced understanding of their impact on a country's happiness levels.
4. Tapping into Social Media: Leveraging social media data enables us to uncover the general sentiment and potentially establish meaningful connections to an individual's happiness.
5. Advancing Regression Models: Exploring innovative multiple regression models empowers us to fine-tune the predictive accuracy, ensuring an optimal fit to the data.
6. Embracing Cultural Aspects: Recognizing the subjective nature of happiness, the inclusion of cultural features allows us to appreciate how diverse perspectives shape individual perceptions of happiness.

4 References

4.1 Literature:

1. Helliwell, J. F., & Wang, S. (2014). The state of world happiness. In World Happiness Report (pp. 1-34). Sustainable Development Solutions Network.

2. Oswald, A. J., & Wu, S. (2010). Objective confirmation of subjective measures of human well-being: Evidence from the USA. Science, 327(5965), 576-579.

3. Stevenson, B., & Wolfers, J. (2008). Economic growth and subjective well-being: Reassessing the Easterlin paradox. Brookings Papers on Economic Activity, 2008(1), 1-87

4.2 Data Sources:

1. https://worldhappiness.report/ed/2023/#appendices-and-data

2. https://www.worldometers.info/world-population/population-by-country/

Consent: I allow sharing of this report with other students in future versions of this class.