**PROJECT 1 REPORT**

**WORLD HAPPINESS**

**Archana Karri**

**Myrna Deseo**

**Neha Sharma**

**Rasika Senadheera**

Submitted as a Requirement for the Data Analytics and Visualisation Course

22nd September 2022

**BACKGROUND**

The year 2022 marks the 10th year anniversary of the World Happiness Report. This report was based on a global survey data of more than 150 countries worldwide, which is called the Gallup World Poll (<https://www.gallup.com/analytics/318875/global-research.aspx>), and reported how people evaluate their own lives. The report can be downloaded online at: <https://happiness-report.s3.amazonaws.com/2022/WHR+22.pdf>.

**METHODOLOGY**

The data was downloaded from the World Happiness 2022 report as an Excel file called “Happiness Data” (<https://worldhappiness.report/ed/2022/#appendices-and-data>). A GitHub repository was created, and each member of our team created branches.

**Data Cleanup**

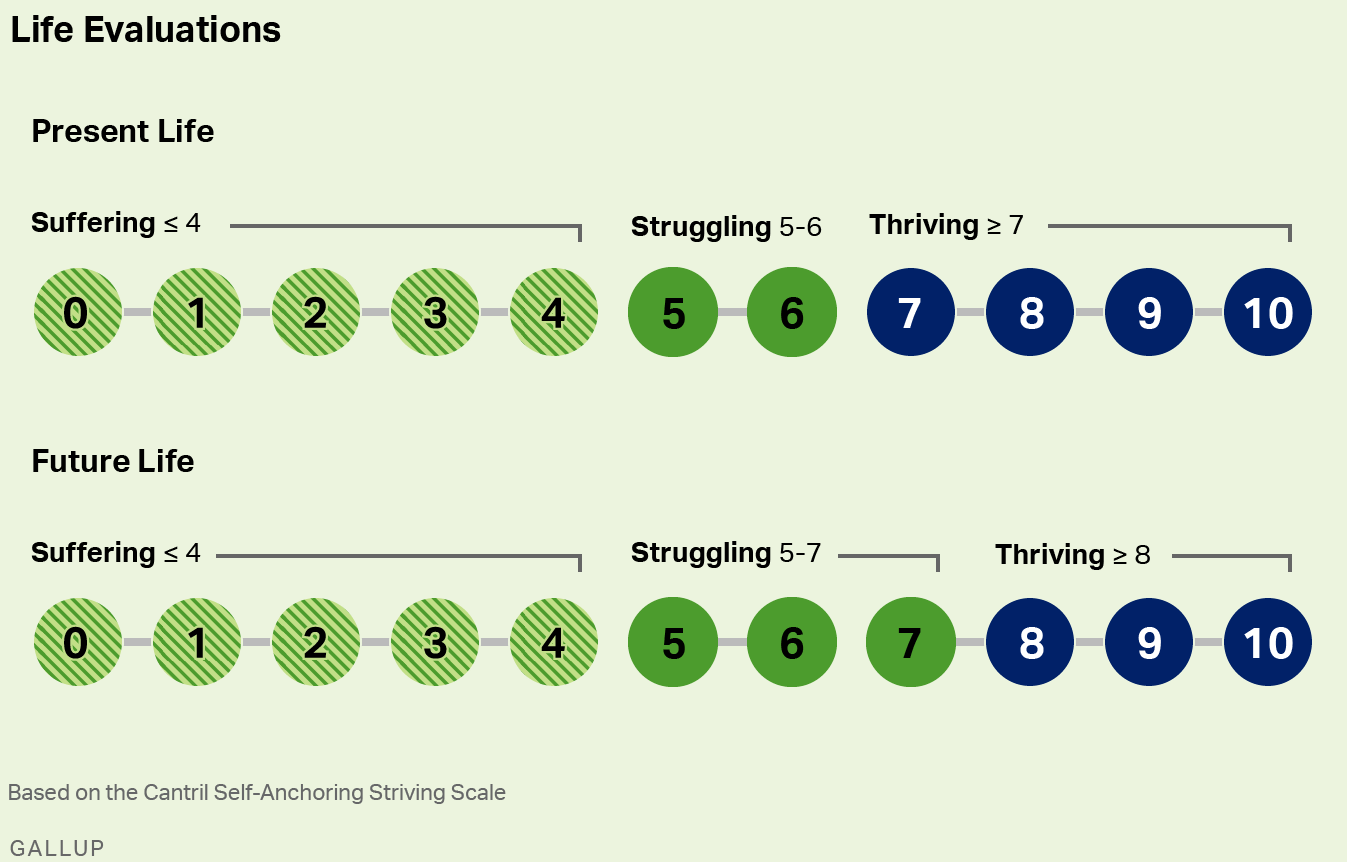
We focused on the data for the year 2021 for all countries. Using Jupyter notebook, we filtered the data to create the year 2021 subset and cleaned up the data by removing cells with null values. The output file from the cleanup can be found in the “FilteredData” folder with file name “Happiness2021.csv.”

From the original Excel file, we also extracted all the available data for Australia and exported it to the “FilteredData” folder with file name “AustraliaData.csv.”

**Data Analysis**

We do not have access to the metadata from the survey, but we have the **2021 summary data from each country surveyed** (Happiness2021.csv). The happiness indicator we used is “Life Ladder” since this is based on the Cantril Ladder, which is a simple visual scale for assessing general life satisfaction, with a scale of 0 to 10. For the survey used in the report, the scaling of “Life Ladder” was based in the image below:

Source: <https://www.gallup.com/analytics/349487/gallup-global-happiness-center.aspx>



From the filtered data, we determined correlations between Life Ladder as the happiness index and the other factors determined in the survey. We plotted the trend of these factors for all the countries and made some observations that were supported by linear regression models.

For the filtered Australia data, we plotted the factors across the years and made some observations on how the trend had changed and which of the factors showed some changes across the years.

**ANALYSIS RESULTS AND OBSERVATIONS**

**Analysis of Trends of the Different Factors Among the Countries (2021 Data)**

Life Ladder was used as the happiness index for this analysis. Among the 100 countries, we determined the top five counties with the highest happiness index and the five bottom countries with low happiness index. The top five countries are Finland, Denmark, Israel, Iceland and Sweden and the five bottom countries with low happiness index are Malawi, India, Zimbabwe, Zambia, and Lebanon (Figure 1).

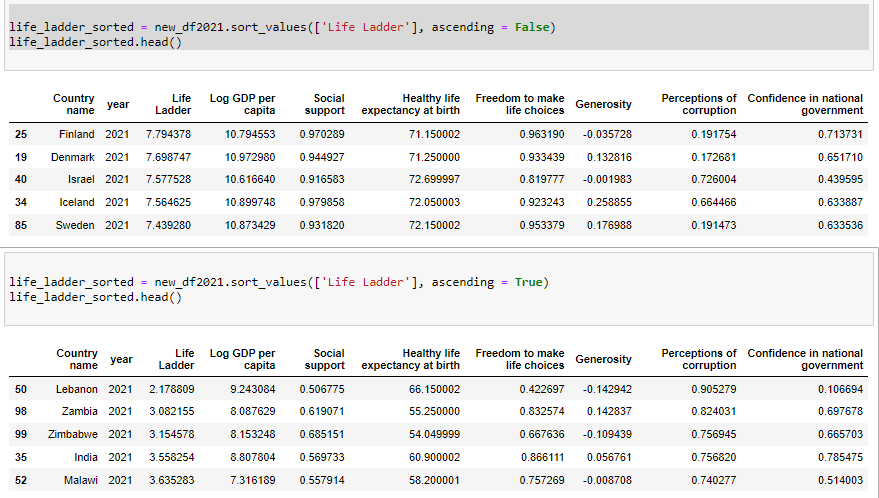


Figure 1. Screenshot of the sorting of countries according to Life Ladder.

Using Life Ladder as the happiness index we created a bar plot of Life Ladder for all the countries. We overlaid the line plot of Generosity to visually determine if there is some trend between happiness and generosity (Figure 2). The plot shows that there is no observable trend between the two parameters across all countries. The regression analysis (Figure 3) confirmed the visual observation whereby R2 gave a value of 0.02, which indicated that there is no correlation between Life Ladder and Generosity.

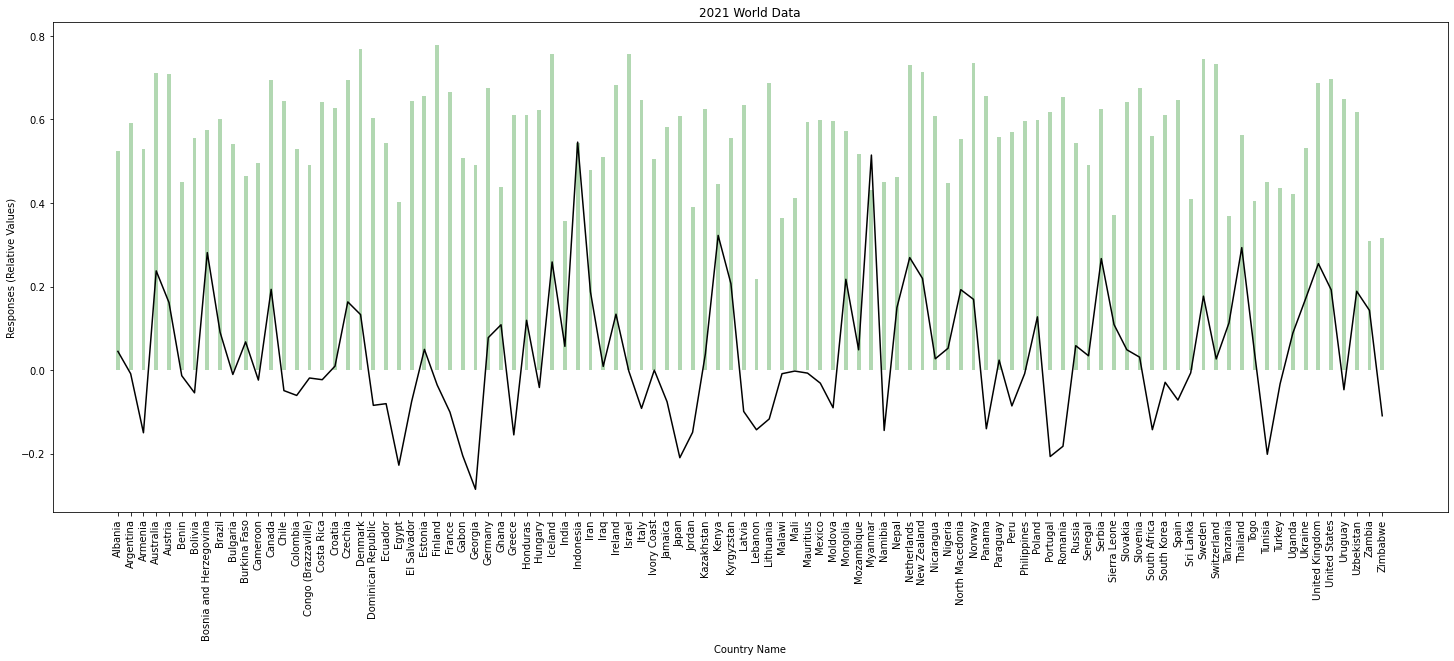


Figure 2. Overlaid plot of Life Ladder and Generosity for all countries.

Chart, scatter chart

Description automatically generated

Figure 3. Scatter plot of Life Ladder vs. Generosity showing the regression line.

The other factors in the data were plotted as line plots and overlaid with Life Ladder to find some correlation among the factors determined in the report. Figure 4 shows the plot of Life Ladder as a bar plot and the overlaid line plots of log GDP per capita, Social support, Healthy life expectancy at birth, Perception of corruption.

To further determine if the individual factors have correlation with Life Ladder, a linear regression analysis was carried out for each factor vs Life Ladder. Figure 5 shows the regression analysis for Life Ladder vs Healthy life expectancy at birth and there seems to be a positive trend or positive relationship between the two although the R2 value is 0.51. This indicates that the model cannot be used to confirm the relationship between Life Ladder and Healthy life expectancy at birth.

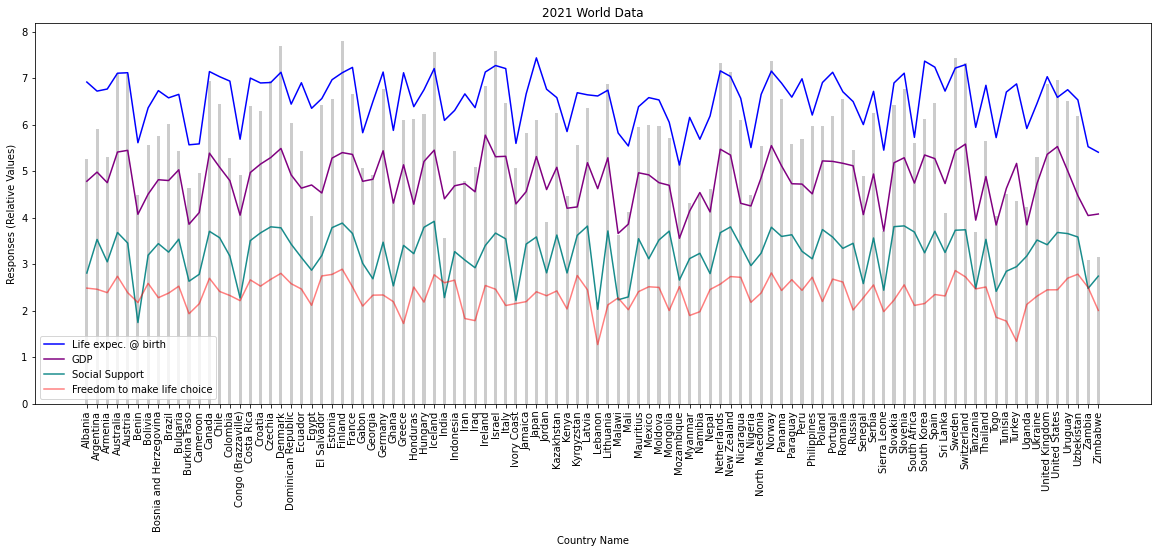


Figure 4. Overlaid plot of Life Ladder for all countries and log GDP per capita, Social support, Healthy life expectancy at birth and Perception of corruption.

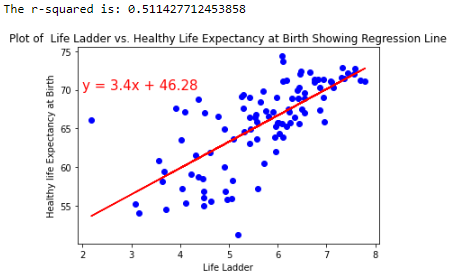


Figure 5. Scatter plot of Life Ladder vs. Healthy life expectancy at birth showing the regression line.

The regression analysis for Life Ladder vs log GDP per capita is shown in Figure 6. Similar to Healthy life expectancy at birth, there is a positive relationship between Life Ladder and log GDP per capita. However, the calculated R2 value of 0.62 suggests that this model cannot also be used to confirm the relationship between Life Ladder and log GDP per capita.

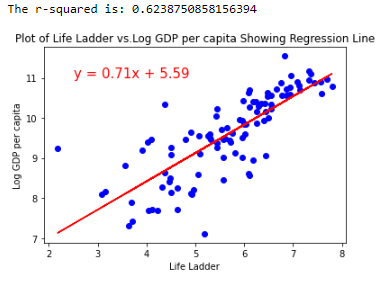


Figure 6. Scatter plot of Life Ladder vs. log GDP per capita showing the regression line.

The scatter plot of Life Ladder vs Social support follows a similar trend as the log GDP per capita. There is a positive relationship that was observed between Life Ladder and Social support as shown in Figure 7, although the calculated R2 value is 0.69. This also suggested that the linear regression model cannot be used to confirm this relationship between Life Ladder and Social support.

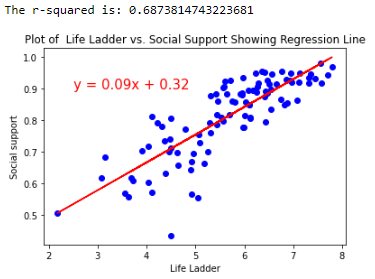


Figure 7. Scatter plot of Life Ladder vs. Social support showing the regression line.

A similar trend was observed in the regressions analysis of Life Ladder vs Freedom to make life choices (Figure 8) whereby it seemed to also follow a positive relationship, but the regression analysis shows a low R2 value of 0.39. This indicates that this model cannot also be used to confirm the relationship between Life Ladder and Freedom to make life choices.

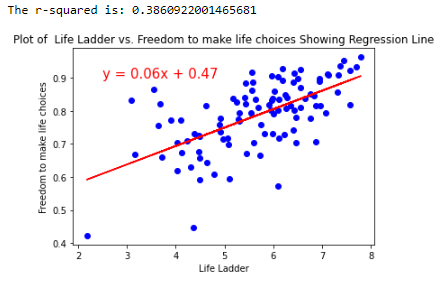


Figure 8. Scatter plot of Life Ladder vs. Social support showing the regression line.

Among the factors plotted in Figure 4, there seems to be a similar observable trend between log GDP per capita and Healthy life expectancy at birth. We did a regression analysis for these factors and the scatter plot in Figure 9 shows a positive relationship where Healthy life expectancy at birth follows an increasing trend as log GDP increases across all countries. The calculated R2 value is 0.75, which is a good indication of the linear relationship but still cannot be confidently used to confirm the relationship between these log GDP per capita and Healthy life expectancy at birth.

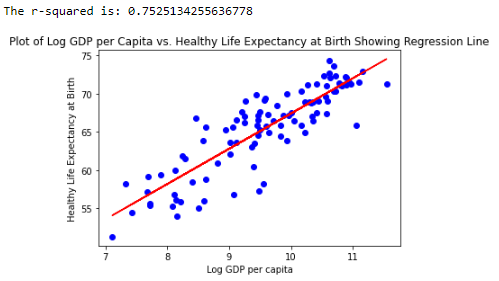


Figure 9. Scatter plot of log GDP per capita vs Healthy life expectancy at birth showing the regression line.

**Analysis of Australia Data**

A subset of the data available for Australia was exported (AustraliaData.csv) and the following factors were plotted for the years 2007 to 2021 (except 2009 where no data was provided):

* Life Ladder
* Social support
* Generosity
* Perceptions of corruption
* Log GDP per capita

The line plot in Figure 10 shows that Life Ladder (purple line), which is the happiness index, has remained relatively consistent across the years as indicated by having almost a straight-line plot. The same was observed for the factors log GDP per capita (purple line) and Social support (blue line). The factors Generosity (green line) and Perceptions of corruption (red line), on the other hand, has observable fluctuations across the years.

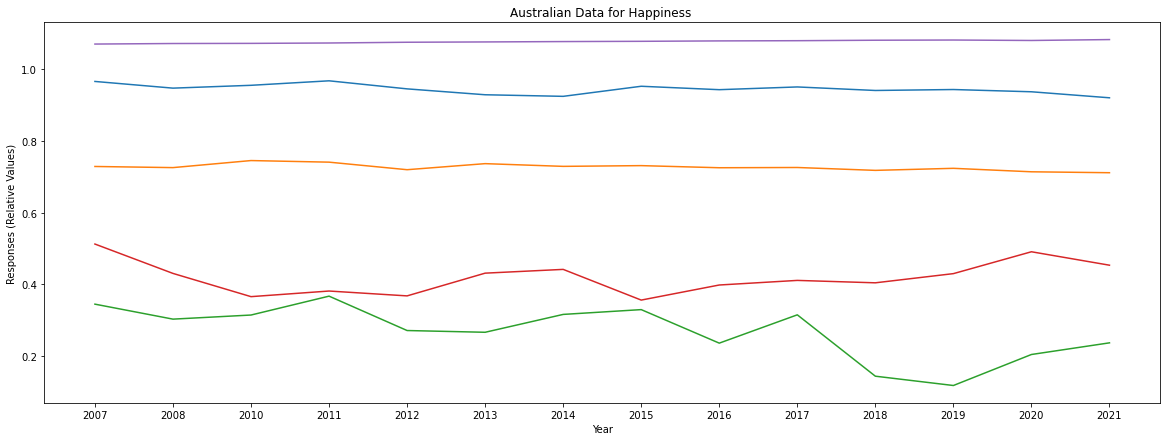


Figure 10. Line plots of Life Ladder (happiness index) and the factors Social support, Generosity, Perceptions of corruption and log GDP per capita the years 2007 to 2021 (except 2009).

The Covid-19 pandemic started in 2020 so we looked closely on the trends from 2018 to 2021. The overlaid line plots in Figure 11 show that Life Ladder (red line) and the factors log GDP per capita, (pink line), Social support (blue line), Healthy Life expectancy at birth (orange line) have not changed much during these years. The trends for Generosity (purple line), Confidence in national government (green line) and Perceptions of corruption (brown line) had some slight change in its trend. From 2019 to 2020, the trend for generosity increased and still maintaining a slight increase until 2021. The trend for Confidence in national government, however, slightly decreased in 2020, but started to increase again in 2021. The inverse is true for Perceptions of corruption; the trend increased in 2020 and decreased in 2021.

Chart, line chart

Description automatically generated

Figure 11. Line plots of Life Ladder (happiness index) and the factors Social support, Generosity, Perceptions of corruption, Healthy Life expectancy at birth and log GDP per capita the years 2018 to 2021.

**REFERENCE**

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2022). World Happiness Report 2022. New York: Sustainable Development Solutions Network.