World Happiness Data Analysis

ThanhVi Dang, Jasmine Coloma, Kyle Ryan
Fordham University, Graduate School of Arts & Sciences

***Abstract* – Exploratory data analysis (EDA) is a powerful tool for understanding and analyzing complex datasets. This study applies EDA techniques to the World Happiness Report, a widely-cited annual survey that ranks 156 countries by their levels of happiness. We review the World Happiness Report from the years 2015 to 2021. The study first reviews the methodology used to compile the World Happiness Report, including the survey questions and scoring system. It then uses EDA methods such as data visualization and statistical modeling to investigate the relationships between different factors and happiness levels. The study results show strong correlations between happiness and factors such as income, social support, healthy life expectancy, and freedom to make life choices. However, the study also reveals some surprising findings, such as the relatively low correlation between happiness and factors such as generosity and perceptions of corruption. Overall, this study demonstrates the usefulness of EDA for gaining insights into complex datasets such as the World Happiness Report. In addition, it provides a rich and nuanced understanding of the factors contributing to happiness and suggests avenues for further research.**

Keywords – Exploratory Data Analysis, Data Mining, Data Visualization, Happiness

I. Introduction

The World Happiness Report began publishing annually in 2012 to coincide with the United Nations' "International Day of Happiness." The report ranks over 150 countries by their happiness levels based on 6 main factors: GDP per capita, social support, health life expectancy, freedom to make life choices, generosity, and perceptions of corruption. The main goal of this analysis is to determine which factors have made a major impact on a country or region's happiness score. This analysis will focus on the happiness score as our target variable. We aim to examine the happiness score and the extent to which any of the main factors have a strong correlation with it.

II. Previous Works Done by Others on the Same Dataset

This dataset, developed by Sustainable Development Solutions Networks (SDSN) to track the world's happiness, has been overwhelmingly popular in the research of human happiness. Since the launch of its first World Happiness Report, it has continually gained worldwide media and academic recognition. This report has been featured in NPR, The Guardian, The New York Times, The Washington Post, U.S. and World Report, Harvard Business Review, Huffington Post, Forbes, etc. Organizations such as the United Nations, along with researchers from various disciplines around the world from top universities use the work of this report to extend our understanding of

happiness and as a measure of this country's success. By far, this is the most downloaded and widely cited data. This has been proven by many trusted media outlets to be an indispensable tool to better understand what makes people happy.

III. Data Sources Details

The World Happiness Report is a survey intended to measure the state of worldwide bliss. The first report was published in 2012 and since then informs policymakers on decisions to help increase overall joy in their country. Currently, the report is completely up to the years 2021 and 2022 and has some data available. The joy scores and rankings utilize information from the Gallup World Survey. The survey utilizes something called the Cantril step which asks the responder to envision the best possible life as being a ten and the worst possible life as being a zero then to place their current life on that scale [2]. The Cantril step covers six different areas: financial generation, social backing, life anticipation, flexibility, nonattendance of debasement, and liberty. The goal of the Gallup World Survey is to assess which nations are currently in a dystopia and then make decisions to raise their quality of life. The file contains the Happiness Score for 153 countries along with the factors used to explain the score [2]. The Happiness Score is a national average of the responses to the main life evaluation question asked in the Gallup World Poll which uses the Cantril step. The Happiness Score can be explained by the following factors: GDP per capita, life expectancy, social support, freedom, generosity, corruption, and residual error.

IV. Results and Interpretation

Time Series Plot

A time series plot is shown below comparing the average happiness score of all countries throughout the years 2015 to 2021. It is evident that the score for happiness has gradually increased over time, despite a slight decrease between 2016 to 2017. In turn, there is a huge jump in the average score for happiness between 2019 to 2021. Although the recent Covid-19 pandemic has caused a lot of unprecedented chaos, it is also possible that it has created a strong impact on newer conceptions of what is important for living a good life as well as collective well-being. A world-wide shutdown is likely what caused many people to rethink their values for quality of life.

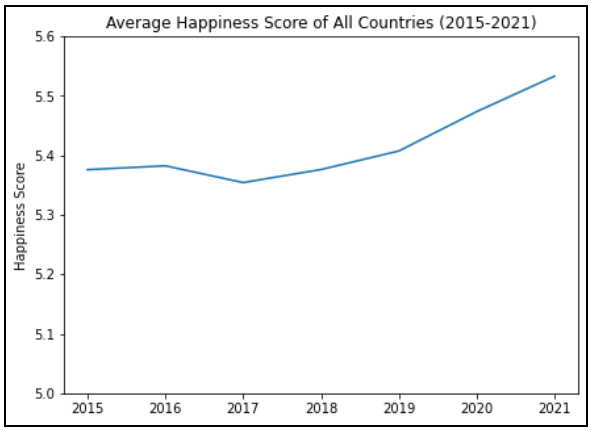


Fig. 1 Average Happiness Score of All Countries (2015-2021)

Scatter Plot

The scatter plot is shown to visualize the relationship between the economy and happiness score. There is a clear correlation between GDP per Capita and national happiness, indicating that a country's income plays a significant role in determining their overall life satisfaction. It

can be interpreted that the increase in one's income leads to the increase in standard of living which does, in turn, create more happiness.

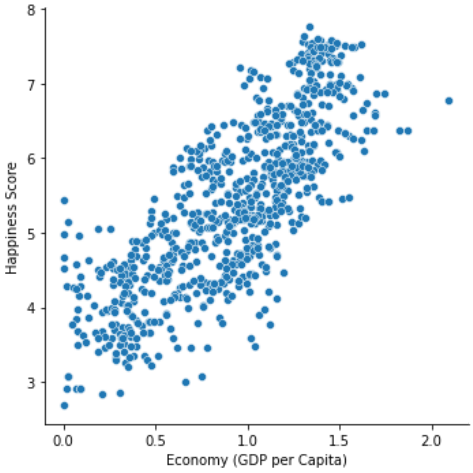


Fig. 2 Scatter Plot of the Happiness Score in Relation to Economy (GDP)

Bar Chart

A bar chart is shown below to display the top ten happiest countries using their overall average happiness score between 2015 to 2021. Switzerland is in the top position, followed by Iceland, Denmark, and Norway. According to the World Happiness Report in 2021, Finland had achieved top position for the fourth year in a row. In turn, Afghanistan returns with the lowest average happiness score, making it the least happiness country out of the group. It is also interesting to note that the United States does not make it in the global top ten.

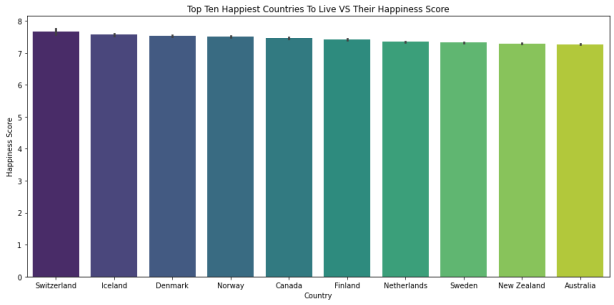


Fig. 3 Top Ten Happiest Countries to Live versus Happiness Score (2015-2021)

Cumulative Plot

Between the 148 countries that this dataset derived from, the breakdown of those countries falls under 10 regions around the world. The majority of the reported countries belong to the Sub-Saharan African region (27%), Central/Eastern Europe takes the second lead (19.5%), and only a small amount in North America (0.01%), as well as Australia/New Zealand (0.01%). As shown in Fig. 4, the cumulative plot shows the average score of each region under the 6 variables we are focusing on. North America is leading with a happiness score of 7.28, whereas Sub-Saharan Africa has a low score of 4.20, which could be due to their low report on GDP per capita and freedom. The countries with the highest average scores tend to be amongst highly developed countries, in comparison to the less developed countries. This, in turn, impacts their perception of freedom, family, generosity, and most importantly, health. Overall, the trends on the plot shows that the Sub-Saharan African region consistently score lower than the rest of the other regions in the world. However, that could be due to their bigger sample size.

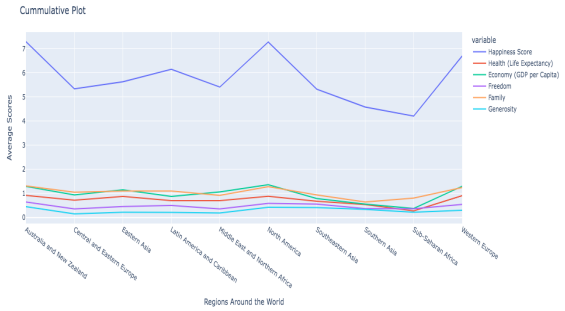
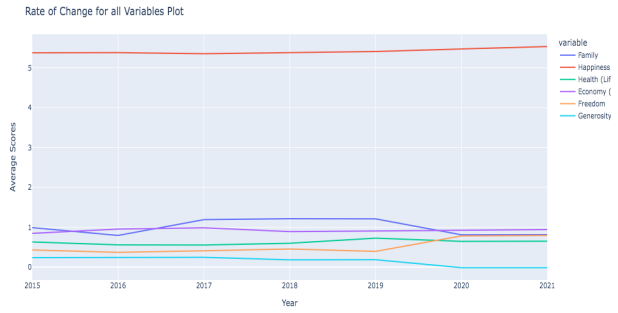


Fig. 4 Cumulative Plot of Average Scores for each Region Around the World.

Rate of Change Plot

This rate of change plot is an extension of Fig.1. It compares all the variables in the data set: *family, happiness score, health, economy, freedom,* and *generosity* between 2015 and 2021. Each line on the plot is in accordance with the variables as it represents the rate of increase or decrease in their mean average score. Based on Fig. 5, the *happiness score* and *family* variables are most consistent throughout the years. Whereas, in *family* and freedom, there are more drastic changes in past years.



**Note: Health (Life Expectancy) values in this graph are divided by 100 to get a smaller value for graph purposes.*

Fig. 5 Rate of Change for all variables (2015-2021)

Looking closely at the rate of change in life expectancy from 2015-2021, Fig. 5 shows an inconsistent trend in health. We see that there is a positive rate of increase from 2017-2019. However, the sudden drop between 2019 and 2020 can take into consideration the COVID-19 pandemic.

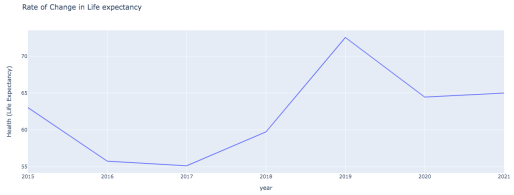


Fig.6 Rate of Change in Health (Life Expectancy) from 2015-2021.

Histogram

As we have seen previously, there is a correlation between happiness and health. To dive deeper into this topic. Fig.7 shows the average happiness score per average health (life expectancy) score. This histogram plot demonstrates that the lower the happiness scores, the lower the life expectancy. The lowest average happiness score (3.91) is associated with a score range of 12.5-17.5 in health, in contrast to a high happiness score of 6.70 with a score range of 102.5 - 107.5 in health. Looking at this, it is important to find ways to improve happiness due to its effect on life expectancy.

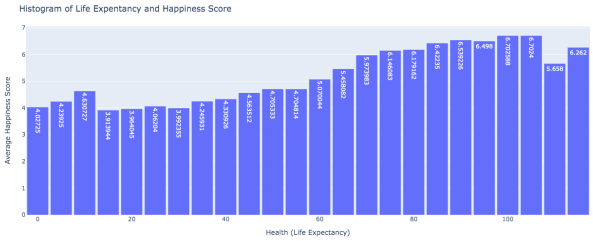


Fig.7 Histogram of Average Happiness Score and Health (Life Expectancy).

Heatmap

A heat map is a graphical representation of data that can be used to visualize the relationship between different variables. Fig. 8 shows that the variables with the highest

correlation to happiness score are health (life expectancy) and economy (GDP per capita). These variables have an intensity value of 0.70 and 0.74, respectively. One possible interpretation of this result is that economic prosperity and good health are essential factors that contribute to happiness. Countries with high GDPs may have more resources and opportunities available to their citizens, which could lead to higher levels of happiness. Similarly, longer life expectancies indicate good health and access to quality healthcare, which also contribute to happiness. However, it is important to note that there may be other factors at play that could influence the relationship between these variables. For example, cultural differences and individual circumstances may also play a role in determining happiness scores. The relationship between these variables may not be as straightforward as it appears, and further analysis may be needed to fully understand the relationship between GDP, life expectancy, and happiness.

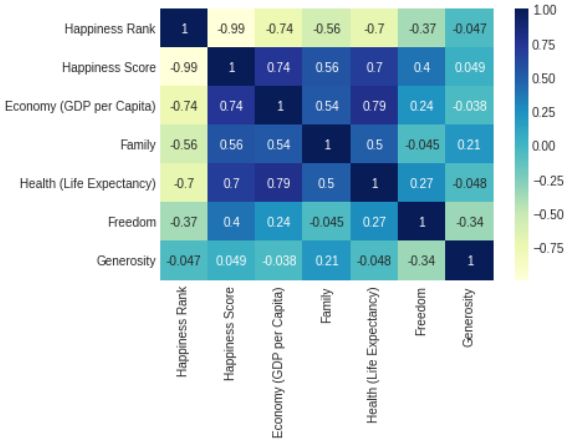


Fig.8 Correlation Heatmap of All Float Columns

Choropleth Map

A choropleth map is a type of map that uses different colors or shades to represent different values or ranges of values. This type of map is often used to display statistical data, such as population density or economic indicators, over a geographical area. In our instance, we are displaying happiness scores over the world map. Each country is shaded according to the value it represents. For example, Fig. 9 shows the happiness score where lighter shades represent happier countries, and dark shades represent countries closer to dystopia. Choropleth maps help visualize patterns and trends in data over a large geographical area. For example, we are able to examine which regions are the happiest at a glance by simply looking at the continent's color. For example, North America and the Nordic countries all are bright shades of yellow, indicating these portions of the world are typically happier. However, it is crucial to be mindful of the limitations of choropleth maps, such as masking important details or variations within regions where there is a high level of aggregation. In our case, where we are only examining countries, we encounter no issues, but if you were to examine regions of each country, then it would be much harder to visualize.

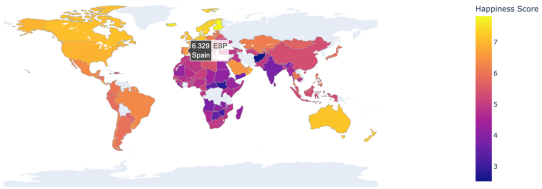


Fig.9 Choropleth Map of Happiness Score

Scatterplot by Region

Now, let's examine a scatter plot colored by region. In this visualization, we plot life expectancy on the x-axis and happiness score on the y-axis in Fig. 10. Each dot's color represents the region the data belongs to, such as North America, Latin America & The Caribbean, Southeastern Asia, and more. Analyzing this graph, you will notice that where the happiness score is the highest, life expectancy is also very high—the region associated with the highest happiness and life expectancy in western Europe. In addition, you may see patterns or trends in the data, such as whether countries with higher happiness scores tend to have longer life expectancies or whether there is a relationship between happiness scores and life expectancy within specific regions. Moreover, as life expectancy decreases, there will be a sharp decrease in happiness scores.

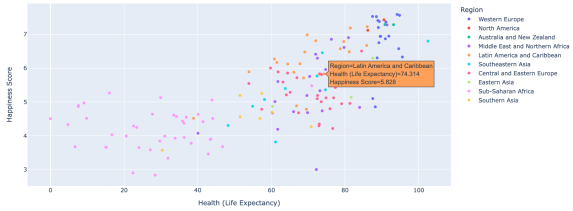


Fig.10 Scatterplot of Happiness Score & Life Expectancy by Region

V. Discussion and Comparisons

Through our analysis, we were able to discover many relationships between happiness and its main factors such as GDP per capita and life expectancy. We found that as variables such as health and the economy increased, so did the overall rank of happiness. Figure 2 shows a clear linear relationship between happiness and GDP, and is shown as well in the correlation heat map in Figure 8. Not only so, but Figure 8 also indicates a correlation between health and happiness, given by the relationship intensity value of 0.7. These findings are significant because it allows us to understand happiness on a macro level. The bar chart and choropleth map (Figure 3 and Figure 9, respectively) indicate the happiest countries based on their mean happiness score between 2015 to 2021. Given these graphs, a strong assumption can be made in which countries with higher levels of GDP and better access to healthcare rank as the top happiest countries, whereas countries with the lowest happiness have lower levels of GDP and less access to healthcare. Furthermore, it can be concluded that significant life events such as the Covid-19 pandemic have played a role in determining life satisfaction. As shown in Figure 6: *The rate of Change in Health (Life Expectancy)*, the sudden drop in change between 2019 to 2020 can be attributed to Covid-19. However, Figure 1 shows that despite such an unprecedented pandemic full of devastation and loss, there is an increase in the mean average happiness score between 2019 and 2020. Ultimately, the nation-wide shutdown has allowed many people to re-evaluate their values in life and overall well-being.

Although the World Happiness dataset is quite comprehensive, it does not take into account other factors such as education, religion, and other cultural complexities. In addition, the dataset can also be perceived as subjective and vague, making it difficult to precisely measure a meaningful comparison between individual people, countries, and regions. Furthermore, the dataset itself

suffered heavily from missing values and inconsistent column names which resulted in heavy data preprocessing. This resulted in omitting certain factors such as government trust, limiting our exploratory data analysis to only 5 out of the 6 main factors.

VI. Conclusion

This study has shown that the economy and health are highly correlated with happiness. Furthermore, our analysis of the World Happiness Report found that countries with higher levels of GDP per capita and longer life expectancy tend to have higher happiness scores. These results align with previous research on the determinants of happiness, which has consistently identified economic and health factors as essential contributors to subjective well-being. Moreover, our study provides further evidence of the strong relationship between these factors and happiness. Therefore, policies that improve the economy and public health also positively impact overall happiness. These findings provide valuable insights for policymakers and researchers and highlight the need for further study of the complex factors that influence happiness. Overall, this study contributes to the growing body of research on the determinants of happiness and provides a valuable perspective on the role of the economy and health in shaping subjective well-being. It also underscores the importance of considering happiness as a key policy goal and the need for continued research into the factors contributing to human happiness.

VII. References

[1] “Home.” Home | The World Happiness Report, https://worldhappiness.report/.

[2] Mathurinache. "World Happiness Report." Kaggle, https://www.kaggle.com/datasets/mathurinache/world-happiness-report. Accessed [date of access].