

# **CSE-591 - DATA ANALYTICS**

## **Project Final Report**

### **World Happiness Report using Apache Spark**

**Ersin Akşar – Feyza Özkefe**

**Işık University, İstanbul – Turkey**

#### **Abstract**

Happiness is one of the most important components in life, but how do we measure it ? For this purpose, the World Happiness Report was created: An annual survey that measures happiness around the globe. It uses six different metrics: (i) healthy life expectancy, (ii) social support, (iii) freedom to make life choices, (iv) Trust (government corruption), (v) GDP per capita, and (vi) generosity. Those metrics combined in an index rank countries by their happiness. In this study, we aimed to show the relationship between these six factors and visualise the world happiness scores that are reported by the Gallup World Poll. This project was made using the Python programming language. We also used the Pyspark library on Python, which includes the Apache Spark framework. Seaborn, Matplotlib and Plotly libraries were used in visualizations. In our analysis, the factors that affect happiness most and the relationships between each other were looked at and all results were visualized

**Keywords:** World Happiness Report, Data Visualisation, Data Analytics, Seaborn, Plotly

#### **1. Introduction**

The happiness of individuals is an important component not only for individuals, but also for the communities, i.e. states, that individuals come together to form. Happy individuals ensure the formation of a happy society so that a better environment can be created. What every good politician and politician aims to achieve is to build a happy society. In this respect, it is important that happiness can be examined numerically. The happiness score is a numerical value of the happiness of the individuals who make up the state, that is, the people living in that country. The happiness score mentioned in the study is a numerical value calculated separately for each state and shows the happiness of the states (society).

#### **2. Methodology**

More than one hundred countries with six metrics to measure happiness are collected from a public database from year 2015 to 2021. The necessary preprocessing on the data has been implemented and merged.

##### **Collected data variables:**

**Country:** Name of the country

**Region:** The region the country belongs to

**Happiness Rank:** The rank of the country based on the Happiness Score

**Happiness Score:** A metric measured based on the happiness scale of 1 to 10.

The English wording of the question is “Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?” This measure is also referred to as Cantril life ladder, or just life ladder in our analysis. (Happiness Score)

**Economy (GDP per Capita):** In purchasing power parity (PPP) at constant 2017 international dollar prices are from the October 14, 2020 update of the World Development Indicators (WDI).

**Family (Social Support):** Social support (or having someone to count on in times of trouble) is the national average of the binary responses (either 0 or 1) to the GWP question “If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?”

**Health (Life Expectancy):** Healthy life expectancies at birth are based on the data extracted from the World Health Organization’s (WHO) Global Health Observatory data repository (Last updated: 2020-09-28).

**Freedom (To Make Life Choices):** Freedom to make life choices is the national average of responses to the GWP question “Are you satisfied or dissatisfied with your freedom to choose what you do with your life?”

**Trust (Government Corruption):** The measure is the national average of the survey responses to two questions in the GWP: “Is corruption widespread throughout the government or not” and “Is corruption widespread within businesses or not?” The overall perception is just the average of the two 0-or-1 responses. In case the perception of government corruption is missing, we use the perception of business corruption as the overall perception. The corruption perception at the national level is just the average response of the overall perception at the individual level.

**Generosity:** Generosity is the residual of regressing national average of response to the GWP question “Have you donated money to a charity in the past month?” on GDP per capita. [1]

### 3. Results

Happiness Score is calculated, with the following six contributing factors, Economy (GDP per Capita), Family, Health, Freedom, Trust and Generosity. Through Fig. 1, the average Happiness Score among 156 countries is 5.40 out of 10. The lowest Happiness Score recorded is at 2.85, which is a country located at Sub-Saharan Africa. Based on the score recording for significant factor contributing to happiness for the year 2019 is health, on average health is rated at 0.72, and the maximum value recorded is 1.14. On average, Family averaged 1.20, meaning most individuals have someone they can trust in times of distress, while Freedom only averaged 0.39; this shows that many people do not have the freedom of choice or think they cannot have the freedom to choose.

**Fig 1.** Summary Statistics for the year 2019 World Happiness Report Results

Variable	Mean	Std Dev	Min	Max
Happiness Score	5.40	1.11	2.85	7.76
Economy	0.90	0.39	0.00	1.68
Family	1.20	0.29	0.00	1.62
Health	0.72	0.24	0.00	1.14
Trust	0.11	0.09	0.00	0.45
Generosity	0.18	0.09	0.00	0.56
Freedom	0,39	0.14	0.00	0.63

**Fig 2.** Summary Statistics for the year 2020 World Happiness Report Results

Variable	Mean	Std Dev	Min	Max
Happiness Score	5.47	1.11	2.50	7.80
Economy	0.86	0.37	0.00	1.53
Family	1.15	0.28	0.00	1.54
Health	0.69	0.25	0.00	1.13
Trust	0.13	0.11	0.00	0.53
Generosity	0.18	0.10	0.00	0.56
Freedom	0.78	0.11	0.39	0.97

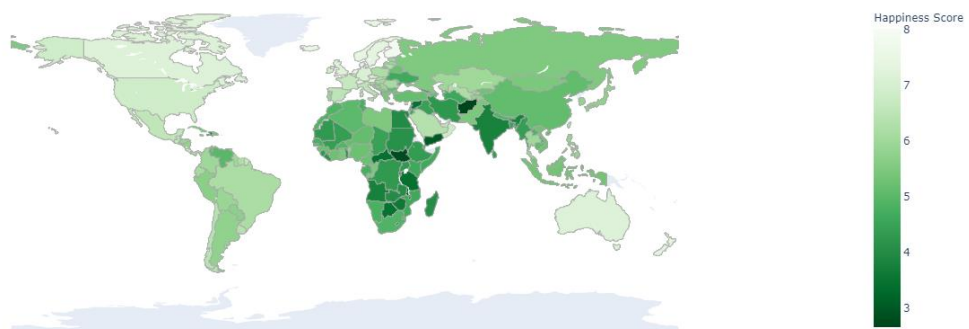
Fig 2. represents the 2020 World Happiness Report summary statistics. Similarly, it covers 7 variables and the 4 statistical summaries. The Average Happiness Score immediately caught attention as it showed that 2019 and 2020 had very similar values at 5.40 - 5.47. However, the lowest Happiness Score fell from last year, with another Sub-Saharan African country recording 2.50. Generosity remains unchanged at an average of 0.18 and Generosity rises slightly to 0.13 respectively, meaning no improvement and stagnant compared to 2019.

**Fig 3.** Summary Statistics for the year 2021 World Happiness Report Results

Variable	Mean	Std Dev	Min	Max
Happiness Score	5.53	1.07	2.52	7.84
Economy	0.97	0.40	0.00	1.75
Family	0.79	0.25	0.00	1.17
Health	0.52	0.21	0.00	0.89
Trust	0.13	0.11	0.00	0.54
Generosity	0.17	0.09	0.00	0.54
Freedom	0.74	0.14	0.26	0.99

Fig 3, represents the 2021 World Happiness Report summary statistics. As per mentioned in the descriptions previously, the summary statistics covers 7 variables and 5 statistical fields. Looking at the mean of Happiness Score which is recorded at 5.53, in comparison to the previous 2 years (2019 and 2020), there has been a increase of 0.06 and the maximum value for Happiness Score has increased to 7.84 recorded by Finland again. As stated previously, Economy, Family, Health, Freedom, Trust and Generosity, all recorded a minimum value of 0.

World Happiness Index 2020



**Fig.4.** Geographical View of World Happiness using Plotly

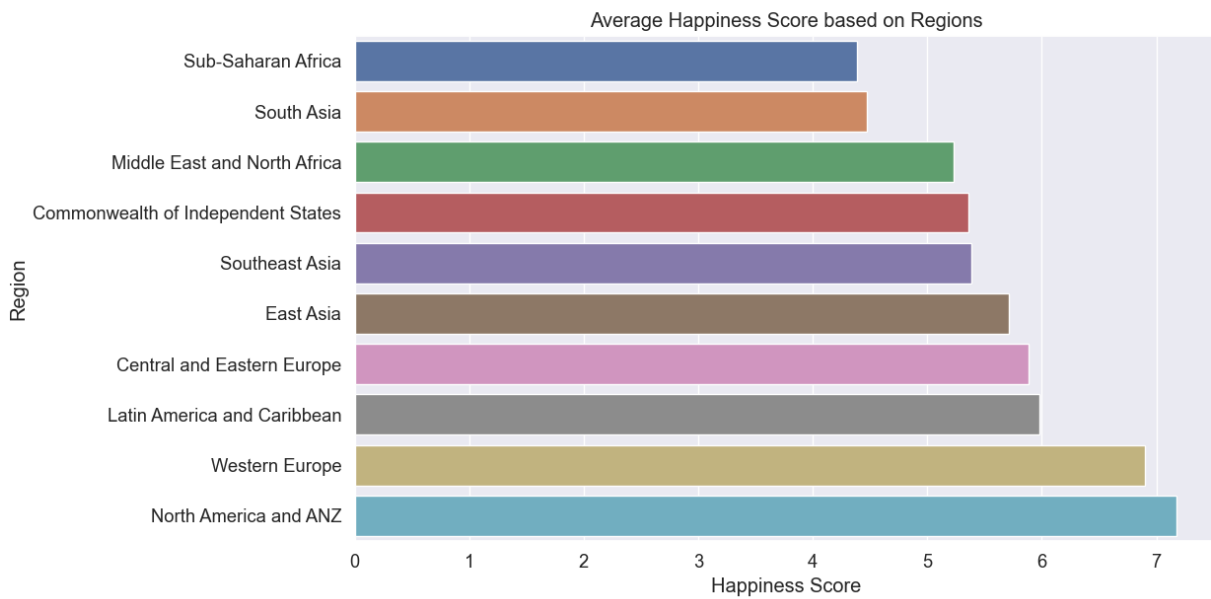
Fig. 4 represents the happiness scores around the world on a map. The colour-coordinated dark green hue represents the lower values; light green hue represents the highest values of happiness of the recorded. Countries with happiest people include Switzerland, Norway, Finland, Denmark, Canada, the United States and Iceland.

Country	Rank	Year
Switzerland	1	2015
Denmark	1	2016
Norway	1	2017
Finland	1	2018
Finland	1	2019
Finland	1	2020
Finland	1	2021

**Fig. 5.** Happiness Rank based on Country (Top 10 happiest for 2015 - 2021)

Fig. 5 shows the ranking of happiness rank according to the 10 happiest countries. Countries consistently ranked among the 20 happiest countries are Switzerland, Denmark, Iceland, Norway, Finland, Canada, Netherlands, New Zealand, Australia, Sweden, Israel, Austria, United States, Costa Rica, Belgium, Ireland and Luxembourg. These are all countries that have emerged consistently over the five years from 2015 to 2021. In addition to these countries that are constantly in the top 20; the remaining countries were constantly shuffling and changed for seven years. As mentioned, some of the happier countries are Switzerland, Norway, Finland, Canada, Iceland, and others that have been portrayed continuously for seven years. The ranking is arranged as followed: #1 to Finland, #2 to Denmark, #3 to Switzerland, #4 to Iceland, #5 to Netherlands, #6 to Norway, #7 to the Sweden, #8 to Luxembourg and #9 to and New Zealand and #10 to Austria

Australia and New Zealand have the most consistent happiness scores from 2015 to 2021. North America firmly lagging behind. Southeast Asia, Central and Eastern Europe, and the Middle East and North Africa regions also do not differ much in average happiness scores; this means that people living in these areas are between average and happy, as their average score suggests. The increases are more pronounced in East Asia, Latin America and the Caribbean, and Western Europe, as a consistent increase in the average happiness score has been demonstrated. Through this bar chart, an assumption can be made to postulate which countries and regions are less happy and happier. It can suggest which areas are above average when it comes to happiness.



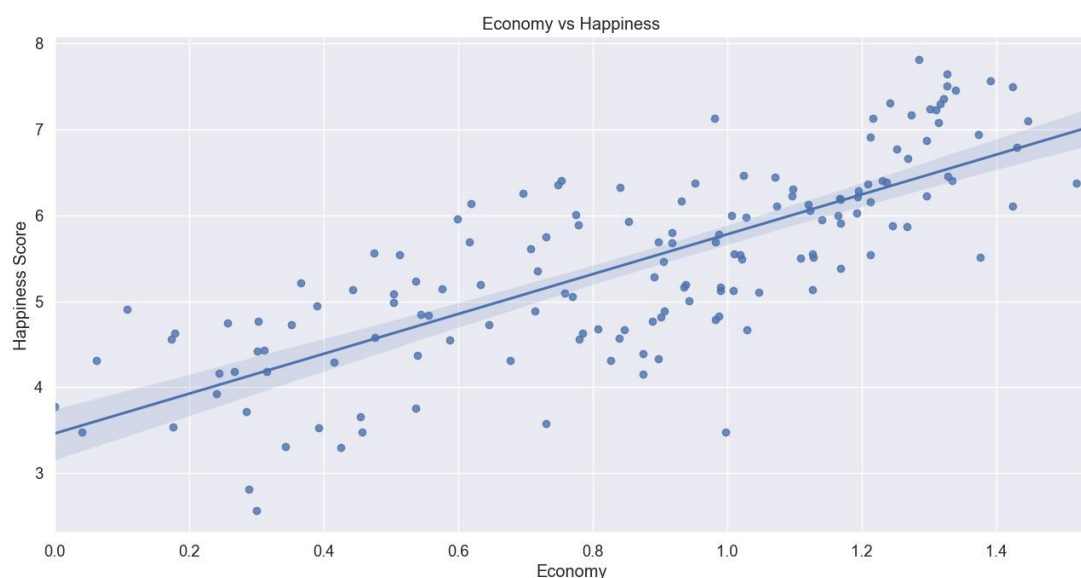
**Fig. 6.** Average Happiness Score based on Regions (Lowest to highest)

Fig. 6. shows the average happiness score based on the regions; Sub-Saharan Africa consists of countries such as Cameroon, Ghana and Kenya, Southern Asia consists of countries, such as India, Nepal and Pakistan, Southeastern Asia consists of countries such as Cambodia, Indonesia and Laos, Central & Eastern Europe consists of countries such as Bulgaria, Croatia, Czech Republic and Estonia., Middle East & Northern Africa consists of countries such as Turkey, Algeria, Bahrain, Dubai, Egypt, Iran and Iraq, Eastern Asia consists of countries such as China, Hong Kong, Japan and South Korea, Latin America & Caribbean consists of countries such as Argentina, Bolivia, Brazil and Mexico, Western Europe consists of countries such as Denmark, Finland, Norway and Switzerland. North America consists of only 2 countries the United States of America and Canada and Australia & New Zealand consists of these 2 countries as mentioned by the region. The lowest average happiness score is at 4.17, which is by the Sub-Saharan Africa region while the highest average happiness score is at 7.304, which is by the Australia and New Zealand region. Surprisingly, the happiest countries in the world are not ranked as the 1st in the average happiness score. This is because, in Western Europe, it consists of many different countries with varying happiness scores, thus affecting the position of this region. [2]



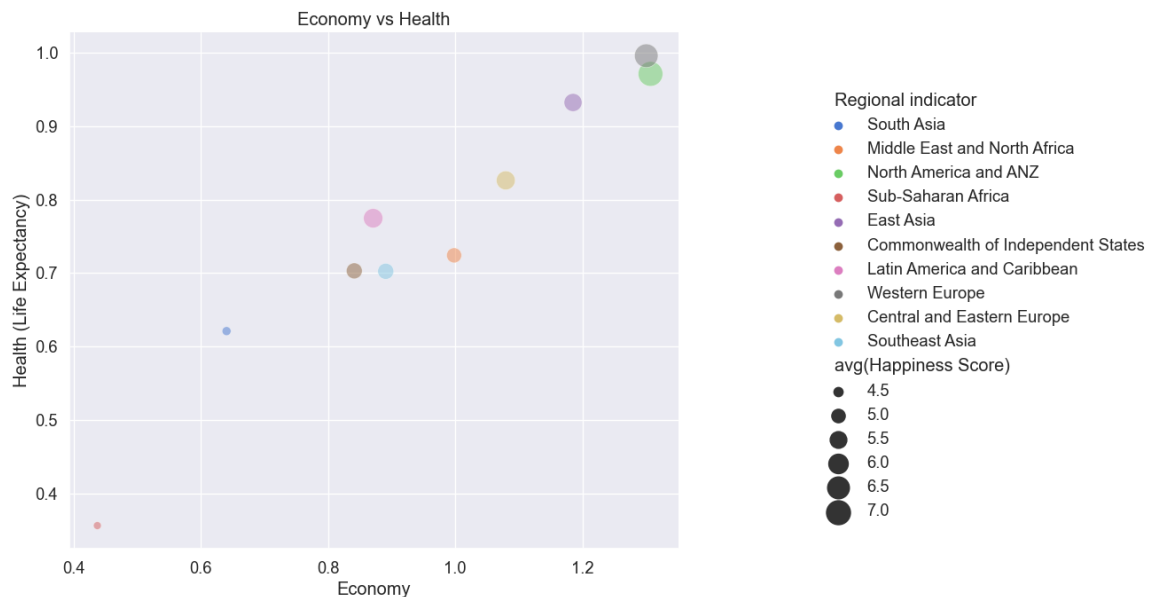
**Fig. 7:** Correlation between 2020 variables

Fig 7 also includes a correlation matrix that examines the relationship of Happiness Score , Economy , Family, Health, Trust Generosity parameters with each other. When you look closely at this matrix, it is seen that it is the economic parameter that shows the highest correlation with the number of happiness. It has a correlation value of 0.78. Then comes health with 0.75. The point here is that the correlation between economy and health is greater than the correlation between the happiness score and the happiness score separately. Therefore, it is understood that the relations between economy and health with each other are more than the happiness score. Based on this, in order for a country's happiness score to be high, not only the economy or not only health scores, but also these two parameters must be high together.



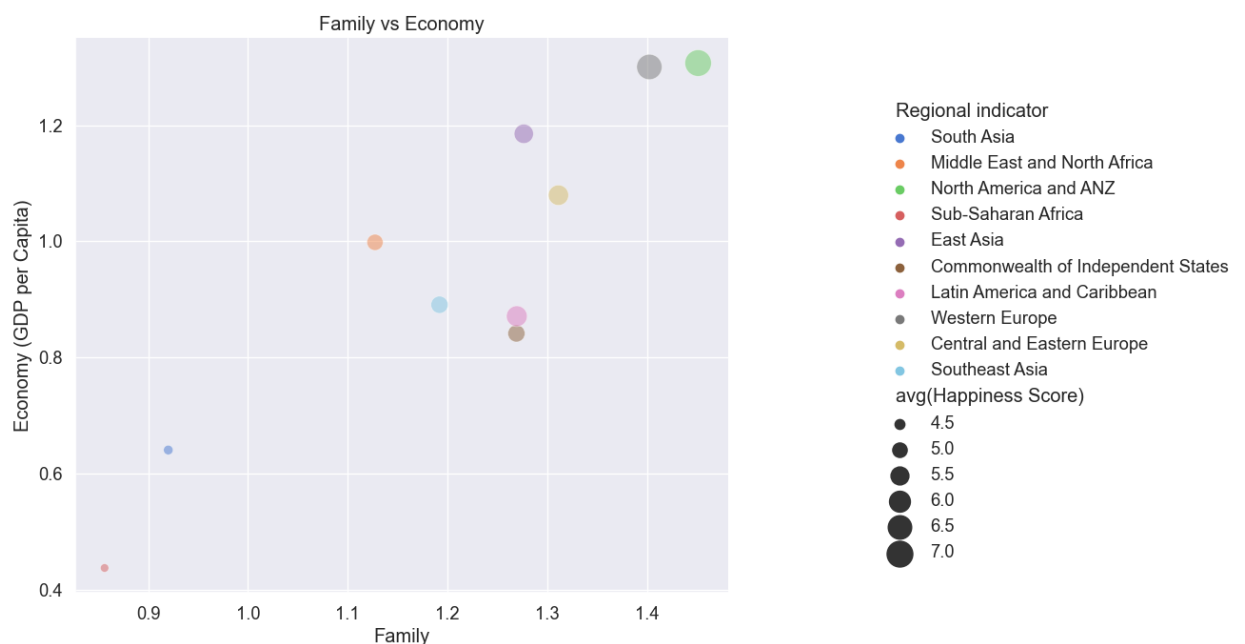
**Fig 8:** Relationship between Happiness Score and Economy

According to the results observed and obtained in Fig. 7, we can look at Fig. 8 to take a closer look at the relationship between economics and happiness. In Fig .8, we can see that the relationship between happiness score and economy is linear. Fig. 8 is drawn using 2020 data.



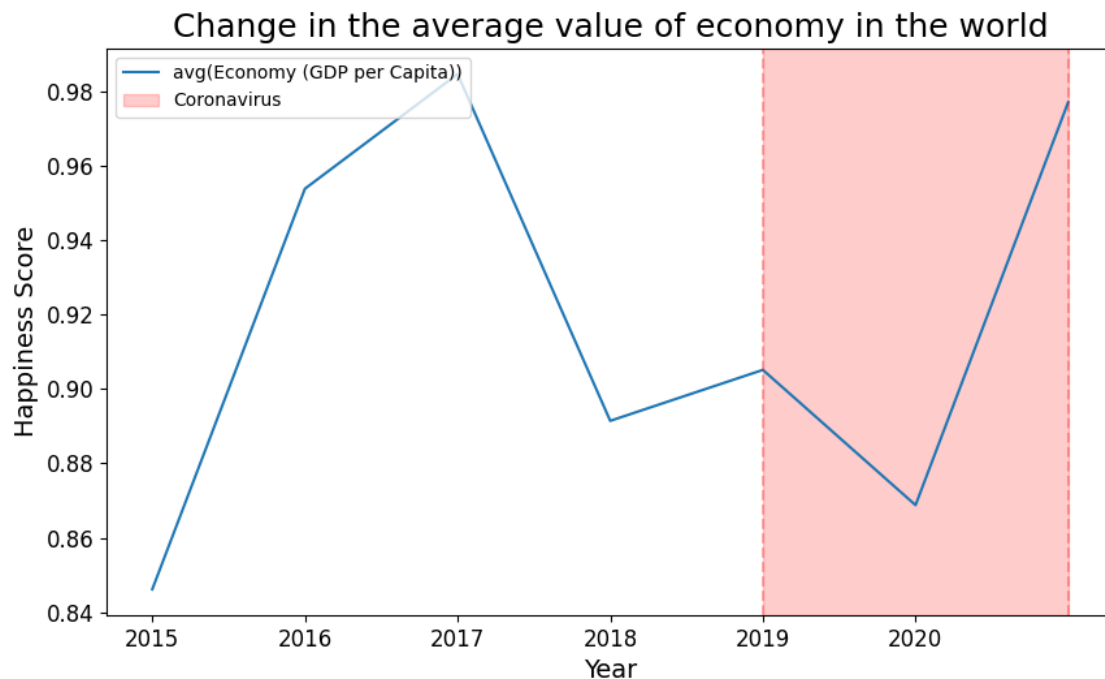
**Fig 9:** Relationship between Health and Economy

In the same way, we can look at Fig. 9 to examine the relationship between health, economy and happiness score. The data used in the drawing of figure 9 is also used for 2020 data to make no difference. As can be seen from the graph, there is a strong relationship between health and economy. . In addition, it has been observed that in countries where these two values are high, living standards and happiness is high in proportion to this.



**Fig 10:** Relationship between Family and Economy

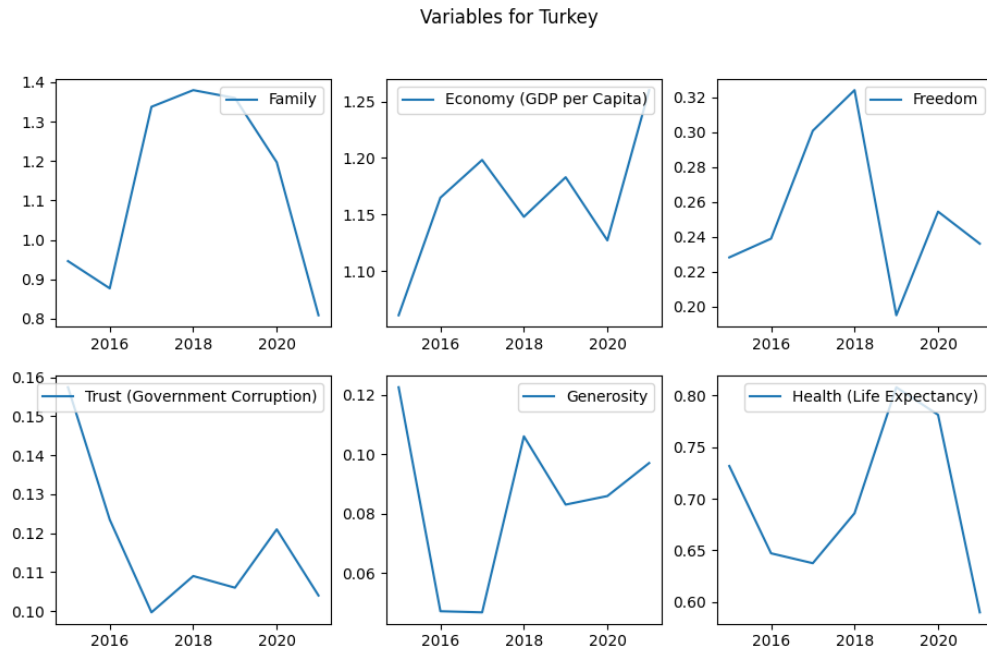
In Fig. 10 we can see How does the Economic Downturn affect families. The decrease in income affects families deeply. Within the nuclear family, stress factors such as job loss, foreclosure or loss of family savings put pressure on parent relationships and the family as a whole. For already low-income families, the shock may be even more severe as basic needs such as food security, health and shelter are not met. In the figure, it is observed that in low-income countries, family scores and happiness are low, on the contrary, in countries where the economy is high, family ties indirectly have high happiness scores. [3]



**Fig 11:** Change in the average value of economy in the world

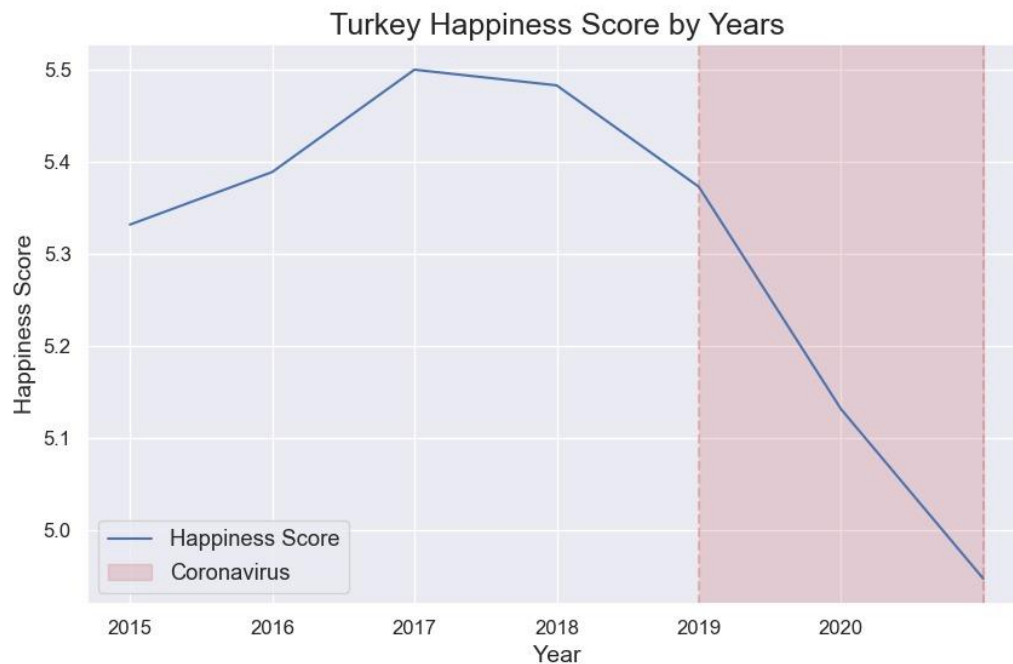
Fig. 11 explains the economy change. The global economy in 2020, after the 1930s Great Depression experienced the deepest recession in economic history. Referred to as the 2020 Great Lockdown in the literature. In the recession, the global economy contracted by 3.5 percent annually in real terms. G4 Central Banks and fiscal policies are extremely important to the situation. A severe financial crisis thanks to their quick and strong reaction. Although the threshold has been reached, the damage in economic activity and unemployment. It is still deep and uncertainty continues. [4]





**Fig 12: Variables for Turkey**

Fig 12 charted the factors affecting Turkey's happiness score by year. When the family factor is examined in this chart, it is observed that it entered 2019 with a serious break towards the end of 2018, and in 2020 this breakage continues and creates a decrease. Again, we can see declines in health, trust, freedom factors. One of the highlights here is the increase in the economic factor. The critical point here is the increase in the economy against the decline in the health factor. This contrast supports our approach, which suggests that economy and health must act together in order to increase the happiness component of a country, as mentioned above. So what we actually expect to happen here is to see Turkey's happiness score drop. Because what we are advocating based on the data in fig4 is that the two most important factors affecting a country's happiness score, namely the economy and health, rise together to increase that country's happiness score. In the light of this information, if we examine Fig 8, we see that Turkey's happiness score has decreased exactly as we expected.



**Fig 13:** Turkey Happiness Score by Years

#### 4. Conclusion

In this study, the graphs of the factors that make up the happiness scores of the countries were drawn and the correlation matrix was prepared in order to better understand the relationship between these factors and the happiness score. The results obtained as a result of the examinations carried out in this direction are the two most important factors that play a role in determining a country's happiness score. These are economic and health factors. When the relationship between these factors and happiness score is examined more closely, it is seen that in order to increase the happiness score of a country, both factors must be increasing harmoniously with each other. In other words, it is not possible to increase the happiness score only by correcting the economy or just health factors. Accordingly, the data of Turkey have been visualized and examined. This data supports the idea put forward. Because while the economic factor in Turkey is improving, the health factor is worsening. When we look at Turkey's happiness score, it seems that the happiness score has fallen as expected. The deterioration of other factors affecting the happiness score in Turkey may have played a positive role in achieving the expected result, but the effects of these factors are not as great as the effects of economic and health factors as mentioned at first on determining the happiness score.

#### 5. References

1. World Happiness Report, <https://worldhappiness.report/>
2. Angela Siew Hoong Lee, Ammar Kudbuddin, Phoey Lee Teh, Phoey Lee Teh, Data Visualisation: World Happiness at a Glance! [https://link.springer.com/chapter/10.1007/978-3-030-34032-2\\_6#citeas](https://link.springer.com/chapter/10.1007/978-3-030-34032-2_6#citeas)
3. What the Economic Downturn Means for Children, Youth and Families <https://www.apa.org/pi/families/resources/economy>
4. 2021 Turkey Economy, Tusiad <https://tusiad.org/>