

COMP1411 Final Project Report

Option3: Evaluating Factors of World Happiness

Group Members:

Gaukhar TURGAMBEKOVA (20065366D),

Yang KE (20100884D).

Motivation

Happiness is the ultimate goal of life for many people. But which aspects of life contribute to happiness the most? Do the answers differ from country to country?

The project is inspired by such questions. Factors like Life expectancy, level of freedom, social support, etc., influence our lives in different ways.

The aim of this project is to rank the factors of happiness for the sake of both governments and citizens. For governments, the ranking of the happiness factors can help them better distribute their fundings to improve people's living environment. For example, if our report shows that good infrastructure is the most important factor of happiness, then the government can invest more in its construction. Additionally, our results could be a piece of advice to policymakers. For instance, if the ranking shows that corruption has the most significant impact on happiness, then the policymakers should punish it with an iron fist. In terms of individuals, our project can assist them in choosing which city or country is the most suitable for residence by comparing the significance of different living factors.

Background

The cornerstone of this project is a report called the World Happiness Report released annually. The team of The World Happiness Report managed to quantify life happiness and its associated factors. The report uses six factors to evaluate the quality of life: GDP, social support, life expectancy, freedom to make life choices, generosity, and perception of corruption. Gallup surveys asking individuals to rate their life based on these 6 factors are carried out worldwide. With the survey results, countries are ranked by comparing them to a hypothetical country called "Dystopia," which has the world's least-happy people and scores 0.

Experts across different fields, including economics, psychology, survey analysis, health, and more, explain how these measurements are justified to evaluate such a subjective feeling as happiness in this report. Besides, the World Happiness Report has gained worldwide recognition as governments and policymakers increasingly take the report into consideration when making decisions.

This project uses the database of the World Happiness Report on Kaggle (<https://www.kaggle.com/ajaypalsinghlo/world-happiness-report-2021?select=world-happiness-report-2021.csv>).

Description

Our main aim is to identify factors that contribute to life quality the most regionally. Correlations are used to measure the contributions of each factor (which is implemented by `cor()` function in R script). The absolute value of a

correlation is proportional to the importance of the factor. By obtaining the factor with the greatest correlation in each region, we are able to draw conclusions of which element is the biggest contributor to happiness in every region.

Implementation

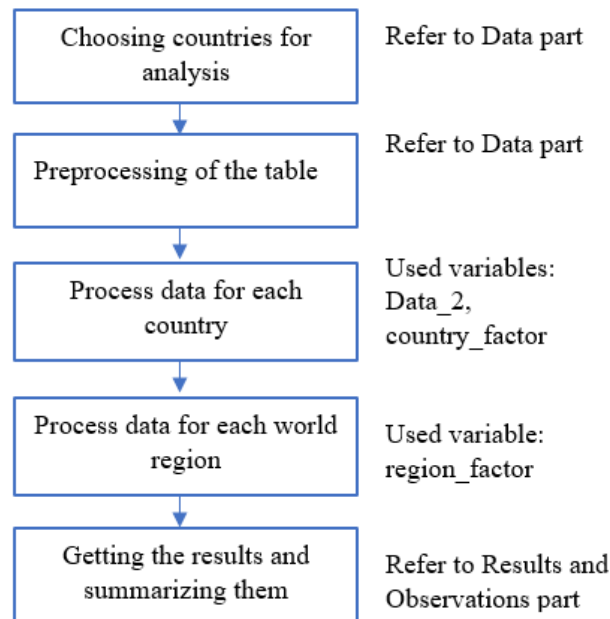


Figure 1. Working steps

In the read-in table, 10 years' data of 6 factors from 40 countries are listed together. The first for-loop is used to process the table by countries, while the for-loop inside it is to process each country by factors. In this loop, the correlation function is used to get measurements of the correlation between “Life Ladder” (i.e. happiness) and the six related factors respectively.

As for the last for-loop, the resulting matrix “country_factor[40][6]” from the previous loop is processed by region. The new matrix “region_factor[10][6]” is obtained by adding up the results of countries from the same region. Besides, “na.rm” function is used to skip missing data.

$$-abs(cor(x,y))^4$$

Figure 2. Correlation Function

Use of Correlation Function:

Correlation is a statistical measure that suggests the level of linear dependence between two variables. Its value ranges from -1 to +1. The negativeness/positiveness only stands for proportional or inverse linear

relationships. The closer the value is to 0, the less relevant the two variables are. Given this feature, the absolute value of correlation is taken to standardise the comparison between different factors.

What's worth mentioning is that when the absolute value is obtained, it's multiplied by the number of population of the country. By doing so, different weights are given to different countries according to their population. Hence, the final results will be more statistically justified.

Variables:

- data_2 - the table consisting of countries information
- country_factor[40][6] - matrix with results per country
row - country; column - factors
- x - life ladder score (column 3 of data_2)
- y - one of the 6 factors (column 4-9 of data_2)
- region_factor[10][6] - matrix with results per each of 10 world parts made from country_factor[40][6] by summarizing values of countries from the same world part.
row - country; column - factors
- factors[6] - storing the names of the 6 factors
- Region[10] - storing the name of the 10 regions
- final[10] - list with final results represented by indexes of the most valuable feature per world part

Data

First of all, the dataset we used is called "world-happiness-report.csv". It's a table download from kaggle

(<https://www.kaggle.com/ajaypalsinghlo/world-happiness-report-2021?select=world-happiness-report-2021.csv>). Its data comes from the World Happiness Report. According to the dataset, 6 main features affecting life happiness are considered and measured:

1. GDP per capita,
2. Social support,
3. Healthy life expectancy,
4. Level of freedom,
5. Generosity,
6. Perception of corruption.

Then, the dataset divides the world into 10 regions:

1. Sub-Saharan Africa
2. South Asia
3. The Middle East and North Africa
4. Southeast Asia
5. Commonwealth of Independent States
6. Central and Western Europe

7. East Asia
8. Latin America and the Caribbean
9. Western Europe
10. North America and ANZ

We take 4 representative countries in each region into consideration in our project. The original table “world-happiness-report” was added with two more necessary columns for this project: “population” and “part of the world” indicating the regions these countries belong to.

The table contains 10-year information per country, so we had more data for analysis. The table is read into the r script as data_2. Below is part of the table

Countryname	year	Life.Ladder	Log.GDP.per.capita	Social.support	Healthy.life.expectancy.at.birth	Freedom.to.make.life.choices	Generosity	Perceptions.of.corruption	Population	part.of.the.world
1 Dominican Republic	2011	5.397	9.475	0.872	64.02	0.848	0.014	0.788	10878246	Sub-Saharan Africa
2 Dominican Republic	2012	4.753	9.489	0.879	64.34	0.840	-0.062	0.727	10878246	Sub-Saharan Africa
3 Dominican Republic	2013	5.016	9.525	0.878	64.46	0.889	0.021	0.752	10878246	Sub-Saharan Africa
4 Dominican Republic	2014	5.387	9.582	0.891	64.68	0.905	-0.020	0.760	10878246	Sub-Saharan Africa
5 Dominican Republic	2015	5.062	9.637	0.893	64.90	0.856	-0.065	0.755	10878246	Sub-Saharan Africa
6 Dominican Republic	2016	5.239	9.691	0.895	65.20	0.873	-0.080	0.737	10878246	Sub-Saharan Africa
7 Dominican Republic	2017	5.605	9.725	0.894	65.50	0.855	-0.121	0.760	10878246	Sub-Saharan Africa
8 Dominican Republic	2018	5.433	9.782	0.862	65.80	0.867	-0.150	0.762	10878246	Sub-Saharan Africa
9 Dominican Republic	2019	6.004	9.821	0.884	66.10	0.877	-0.123	0.746	10878246	Sub-Saharan Africa
10 Dominican Republic	2020	5.168	9.802	0.806	66.40	0.835	-0.128	0.636	10878246	Sub-Saharan Africa
11 South Africa	2011	4.931	9.447	0.858	51.46	0.835	-0.154	0.819	59622350	Sub-Saharan Africa
12 South Africa	2012	5.134	9.453	0.907	52.42	0.590	-0.163	0.838	59622350	Sub-Saharan Africa
13 South Africa	2013	3.661	9.461	0.839	53.38	0.714	-0.077	0.800	59622350	Sub-Saharan Africa
14 South Africa	2014	4.828	9.464	0.881	54.34	0.794	-0.117	0.820	59622350	Sub-Saharan Africa
15 South Africa	2015	4.887	9.460	0.898	55.30	0.862	-0.127	0.853	59622350	Sub-Saharan Africa
16 South Africa	2016	4.770	9.450	0.875	55.70	0.774	-0.070	0.813	59622350	Sub-Saharan Africa
17 South Africa	2017	4.514	9.450	0.870	56.10	0.787	-0.129	0.865	59622350	Sub-Saharan Africa
18 South Africa	2018	4.884	9.444	0.841	56.50	0.753	-0.050	0.841	59622350	Sub-Saharan Africa
19 South Africa	2019	5.035	9.432	0.848	56.90	0.738	-0.134	0.820	59622350	Sub-Saharan Africa
20 South Africa	2020	4.947	9.332	0.891	57.30	0.757	-0.015	0.912	59622350	Sub-Saharan Africa
21 Tanzania	2011	4.074	7.649	0.883	53.04	0.736	-0.046	0.816	56313438	Sub-Saharan Africa
22 Tanzania	2012	4.007	7.663	0.832	53.78	0.577	0.213	0.887	56313438	Sub-Saharan Africa
23 Tanzania	2013	3.852	7.699	0.803	54.52	0.654	0.054	0.859	56313438	Sub-Saharan Africa
24 Tanzania	2014	3.483	7.734	0.788	55.26	0.654	0.110	0.878	56313438	Sub-Saharan Africa
25 Tanzania	2015	3.661	7.764	0.790	56.00	0.759	0.149	0.906	56313438	Sub-Saharan Africa
26 Tanzania	2016	2.903	7.800	0.838	56.50	0.775	0.179	0.739	56313438	Sub-Saharan Africa
27 Tanzania	2017	3.347	7.836	0.705	57.00	0.800	0.116	0.654	56313438	Sub-Saharan Africa
28 Tanzania	2018	3.445	7.859	0.875	57.50	0.807	0.153	0.612	56313438	Sub-Saharan Africa
29 Tanzania	2019	3.640	7.886	0.887	58.00	0.850	0.100	0.589	56313438	Sub-Saharan Africa
30 Tanzania	2020	3.786	7.881	0.740	58.40	0.830	0.295	0.521	56313438	Sub-Saharan Africa

Figure 2. Part of an edited table

After the first correlation function, country_factor[][]

3).

	V1	V2	V3	V4	V5	V6	V7	V8
1	5.397	9.475	0.872	64.02	0.848	0.014	0.788	10878246
2	4.753	9.489	0.879	64.34	0.840	-0.062	0.727	10878246
3	5.016	9.525	0.878	64.46	0.889	0.021	0.752	10878246
4	5.387	9.582	0.891	64.68	0.905	-0.020	0.760	10878246
5	5.062	9.637	0.893	64.90	0.856	-0.065	0.755	10878246
6	5.239	9.691	0.895	65.20	0.873	-0.080	0.737	10878246
7	5.605	9.725	0.894	65.50	0.855	-0.121	0.760	10878246
8	5.433	9.782	0.862	65.80	0.867	-0.150	0.762	10878246
9	6.004	9.821	0.884	66.10	0.877	-0.123	0.746	10878246
10	5.168	9.802	0.806	66.40	0.835	-0.128	0.636	10878246
11	4.931	9.447	0.858	51.46	0.835	-0.154	0.819	59622350
12	5.134	9.453	0.907	52.42	0.590	-0.163	0.838	59622350
13	3.661	9.461	0.839	53.38	0.714	-0.077	0.800	59622350
14	4.828	9.464	0.881	54.34	0.794	-0.117	0.820	59622350
15	4.887	9.460	0.898	55.30	0.862	-0.127	0.853	59622350
16	4.770	9.450	0.875	55.70	0.774	-0.070	0.813	59622350
17	4.514	9.450	0.870	56.10	0.787	-0.129	0.865	59622350
18	4.884	9.444	0.841	56.50	0.753	-0.050	0.841	59622350
19	5.035	9.432	0.848	56.90	0.738	-0.134	0.820	59622350
20	4.947	9.332	0.891	57.30	0.757	-0.015	0.912	59622350
21	4.074	7.649	0.883	53.04	0.736	-0.046	0.816	56313438
22	4.007	7.663	0.832	53.78	0.577	0.213	0.887	56313438
23	3.852	7.699	0.803	54.52	0.654	0.054	0.859	56313438
24	3.483	7.734	0.788	55.26	0.654	0.110	0.878	56313438

application of the we get the matrix (refer to Figure

Figure 3. country_factor matrix

	V1	V2	V3	V4	V5	V6
1	50663482	89805768	56107465	33144045	39383252	42606812
2	1296824112	516955904	1177457369	898906067	679451768	988417879
3	35485574	45147492	47404270	78045769	76725822	98523423
4	144239532	263441074	141961764	199263573	164511213	114861959
5	53779407	127474561	36907566	53846162	30176466	35385157
6	58663546	26124045	56214459	19486922	31085487	51710815
7	816162266	217130701	NA	379905042	672120564	NA
8	230712758	183218768	268006705	49524196	48395317	178772292
9	47036002	92675051	83594064	116206906	22523959	94582050
10	217533457	188516215	172264248	124472172	191329956	39660381

Figure 4. region_factor matrix

Summarizing information for each world part, we get our final results with coefficients for each feature.

Results and Observations

	Region	Biggest.Factor
1	Sub-Saharan Africa	Social support
2	South Asia	Log GDP per capita
3	Middle East and North Africa	Perceptions of corruption
4	Southeast Asia	social support
5	Commonwealth of Independent States	social support
6	Central and Eastern Europe	Log GDP per capita
7	East Asia	Healthy life expectancy at birth
8	Latin America and Caribbean	Healthy life expectancy at birth
9	Western Europe	Freedom to make life choices
10	North America and ANZ	Log GDP per capita

Figure 5. Results per world parts

As it is seen in the resulting table (refer to Figure 5), Social support and Log GDP per capita repeat three times, Healthy expectancy at birth - 2 times, and Freedom to make life choices and Perception of corruption only once each. As a result, these socio-economic characteristics can be considered to be the most contributing factors to people's happiness rate.

Discussions

The report results, however, are not completely precise. There are multiple reasons for this. The first reason is the absence of the data of some countries. Besides, outdated population numbers are taken from open sources, which cannot

be completely precise as the world population rapidly grows. Additionally, the 6 factors are not fully independent from each other, let alone the limited reliableness of the measurements against happiness. Considering all these limitations, the research results can be further improved by active cooperation with official public information sources.

Regarding the application of this knowledge in practice, this research concludes that the essential characteristics contributing to people's lives are social benefits and financial support from the government. That means that countries with higher financial freedom provided by governmental social support tend to have happier citizens and, thus, a higher happiness rate. The happiness rate research is crucial for government social engineering specialists to create better life and work conditions for their citizens. The importance of this research is not limited to government workers. As it defines people's most desired needs, it also can be applied by politicians in their election campaigns and by electors themselves to know which candidate better satisfies their requirement.