CSE 519 Fall 2020: Data Science

Course Project Final Report

Date:12/06/2020

Topic: What factors make people happy in a country?

1.Background

Happiness is the ultimate goal of everyone pursuing a better life. From a higher perspective, it has increasingly been considered as a proper factor of social progress and a goal of social policy, thus more and more research has been conducted on this topic to find what factors can affect human being's happiness.

The World Happiness Report is a survey of global happiness based on respondent ratings of their own lives. The happiness scores for 138 countries are collected from Gallup World Poll. Each responder will answer the Cantril Ladder question based on their self-evaluations, using a scale from 0-10 to rate how they feel about their current satisfactions of lives. Then the researchers use six factors to explain how they affect the happiness scores. The six factors are real GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption [1]. In 2020, Finland, Denmark, Switzerland are the happiest countries among 138 countries.

2.Preliminary work

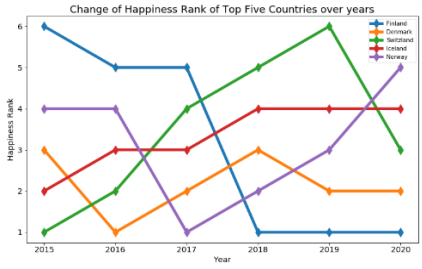
Many researches have focused on the human satisfaction of lives. Researchers have developed different questionnaires to measure personal happiness throughout the world. The World Value Survey provides cross-country data on self-reported life satisfaction [2]. The Oxford Happiness Questionnaire (OHQ) has also been used widely to estimate personal happiness, considering three components: frequency and intensity of positive affect, average level of satisfaction, and absence of negative feelings [3]. The Subjective Happiness Scale (SHS) uses a simple four-question structure to measure respondents' subjective happiness [4].

Our aim for this project is to look deeper into this report data to objectively measure happiness, and eventually to explore the most effective factors that affect happiness.

3. Exploratory analysis

3.1 Answers to basic questions from data

3.1.1. Where are the top ranked countries located and how stable are their rankings?



Top 5 countries in 2020 are all from Nordic:

1.Finland 7.81 out of 10 in happiness score

2.Denmark 7.65

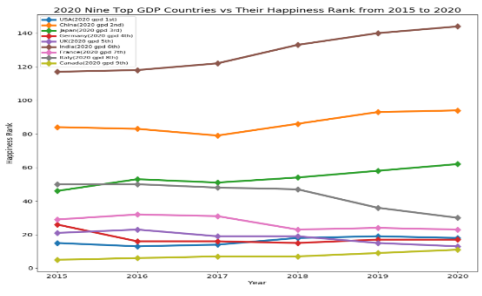
3.Switzerland 7.56

4.Iceland 7.50

5.Norway 7.49

Their ranks have been considerably stable and are always within first six places among all the countries from 2015 to 2020.

3.1.2. Where are the world's largest economies ranked?



In 2020, the nine top GDP countries in sequence are: USA (18th) in happiness ranking, China(94th), Japan(62nd), Germany(17th), UK(13rd), India(144th), France(23rd), Italy(30th), and Canada(11st).

India ranked 6th place in total GDP; however, its happiness ranking in the World Happiness Report is around 144th. The reason behind its low ranking is ascribed to the low GDP per capita as opposed to high national overall GDP. The same reason applies to the situation in China. India and China have the world's largest population.

3.1.3 Which factors are important in measuring happiness according to the report?

Table 20: Decomposing the happiness difference between a hypothetical average country and Dystopia

	Average country	Dystopia	Explained excess happiness over Dystopia due to	Share of explained excess happiness over Dystopia due to
Happiness	5.47	1.97		
Logged GDP per capita	9.3	6.49	.87	.25
Social support	.81	.32	1.16	.33
Healthy life expectancy	64.45	45.2	.69	.2
Freedom to make life choices	.78	.4	.46	.13
Generosity	-.01	-.3	.19	.05
Perceptions of corruption	.73	.94	.13	.04
Sum of explained excess over Dystopia			3.5	1

The six factors, Logged GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity and Perceptions of corruption explain three-quarters of the variation in national annual average happiness scores among all countries, using data from the years 2005 to 2019 [1].

According to the report, a pooled OLS regression is used to explain the average happiness score across countries. Social support accounts for 33% of the total score, which

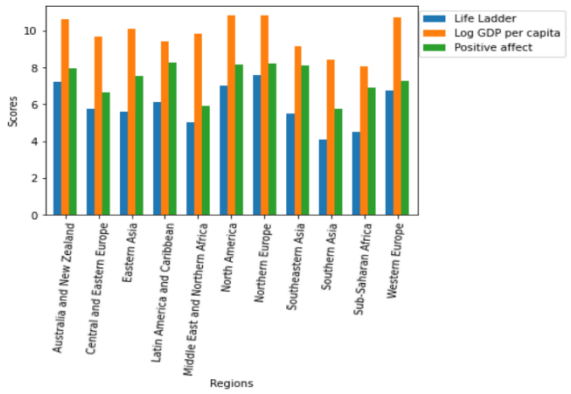
becomes the single largest contributing factor, followed by GDP per capita (25%) and healthy life expectancy (20%) [1].

3.2 Regional Analysis

3.2.1 Data pre-processing

Missing values: Due to the discontinuation or lack of dataset, we fill the missing values by the column means. Also, we impute gaps between years by applying linear interpolation. We do so to ensure smooth year-to-year estimates based on current and historical data and by assuming linear change.

3.2.2 Latin America and Caribbean have low GDP per capita, but why do they have high happiness scores?



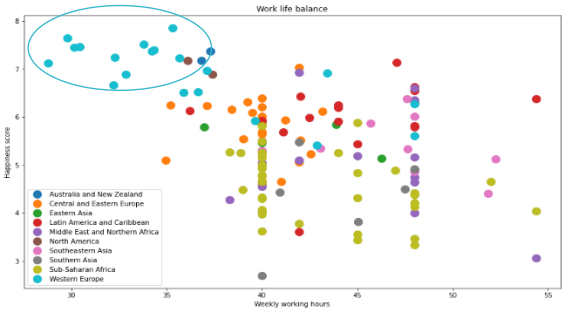
Some Latin American countries have a higher happiness score than developed countries from Western Europe. What is it about the region that makes its people so happy? Money can't buy happiness, and this is clearly reflected by this group of population. One of the reasons behind this is that they have an invaluable asset, a high positive life attitude or positive emotion. This all gets explained in the Positive Affect, a factor in the World Happiness Report recording the score of laughter and enjoyment feelings during a lot of yesterday. From the bar chart we can see that Latin American and Caribbean countries score highest in Positive Affect in green color.

People in Latin America are highly interconnected. Even though they experience low GDP growth and high levels of corruption and crime, they are extraordinarily bound to each other, and braced by friendship and family. The culture unites people and sparks happiness.

3.2.3 Nordic has highest happiness score, do they also have high suicide rates?

3.2.3.1 Why Nordic countries are the happiest?

Nordic countries consistently rank the top in the happiness report, which suggests Nordic people are the happiest around the world. We first assumed that this is because the Nordic countries have better social support, but after we make comparisons among other regions, Nordic countries do not show a significantly higher trend in terms of social support. This led us to find other explanations.



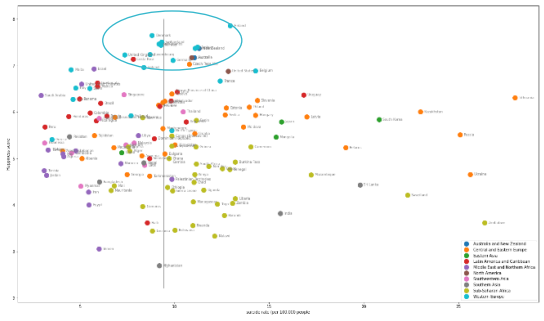
(X: Weekly working hours; Y: Happiness Score)

We found that a good work life balance could be a reason for this phenomenon. According to the research from Agosti (2017) [8], work life balance shows a significant correlation with a positive time experience in private life and well-being. We first chose data of weekly working hours throughout the world in order to find out whether lower working hours contribute more to happiness.

From the scatter plot we can see that the Western European countries, especially the Nordic countries (dots with the blue colors), have significantly fewer working hours than the other regions do. While the normal working hours for most countries are 40 and 48 hours per week, Nordic people work only 30 to 35 hours per week. It is reasonable that with fewer time spent at work, people will have more time to be with their families and friends, or to explore their hobbies and further enjoy their life. Hence their life will be relatively happier than that of people from another region of the world.

3.2.3.2 The suicide paradox

While we can see that the Nordic countries rank top in happiness which makes people think there should be less people commit suicide intuitively, a common stereotype about the Nordic countries is that, as they are plunged into darkness half the year, depression is rife and suicide rates get higher.



(X: suicide deaths/100,000 people;
Y: Happiness Score;
Vertical Line: median of suicide deaths = 9.43)

We then found the worldwide suicide rates, which suggests the number of deaths per 100,000 people. From the plot we can see that the suicide rates in Nordic Countries (with the light blue colors) do show some reverse patterns compared to their top rankings in happiness. They have a suicide rate above the median of 9.43.

The research conducted by Ejrnæs & Greve (2017) [9] may offer an explanation for this. They suggest that people's perception of their position in society is strongly correlated with their level of happiness. An individual's subjective position in society is a more important predictor of happiness than the common objective measurements such as income, education and labor market position. It might be more difficult to be unhappy in an otherwise happy society because it creates a stronger contrast to how you are feeling if you are surrounded by very happy people.

3.3 GDP per capita Analysis-Three Level GDP

Despite the fact that economic factors cannot explain all about happiness, it always plays an important role. Next, we want to investigate how different factors perform on countries with different GDP levels. We divide the countries into 3 groups based on their quartile positions of GDP per capita. The high-level group are the first 33% countries with the highest GDP. Then we calculate the correlations of difference factors with their corresponding happiness scores in each GDP group. From the result, we found some interesting patterns.

1. Some factors show a high correlation with happiness in the high GDP level countries, while a low correlation in the low GDP level countries.

For example, corruption perception has a coefficient of 0.78 in high GDP level countries, 0.22 in mid GDP level countries, and only 0.0041 in low GDP level countries. Similar patterns observed on voice and accountability, freedom of expression, average annual working hours, etc. This can be explained because developed countries have a higher emphasis on their freedom, rights and the work-life balance.

	Indicators	High Gdp Level	Mid Gdp Level	Low Gdp Level
8	corruption perception	0.780	0.220	0.041
21	freedom of expression	0.540	0.110	0.050
25	average annual hours worked	0.550	0.150	0.070

2. On the contrary, some factors show a relatively high correlation with happiness only in low GDP level countries. For example, mobile telephone subscriptions (per 100 people) have a coefficient of 0.41 in the low GDP countries, while having a negligible correlation in the rest two groups. Other factors like child mortality rate, water and sanitation also show increased coefficients from high level to low level countries.

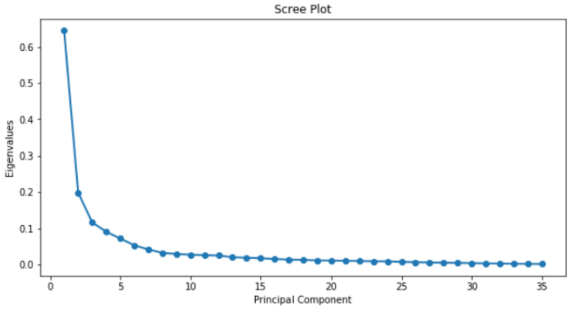
	Indicators	High Gdp Level	Mid Gdp Level	Low Gdp Level
27	mobile telephone subscription	0.037	0.005	0.410
13	water and sanitation	0.130	0.160	0.550
14	child mortality rate	-0.130	-0.190	-0.450

For economically underdeveloped areas, factors that can represent people's basic needs will be more relevant to people's happiness.

In conclusion, we found that in high GDP level countries, factors that represent the advanced needs will be more relevant to happiness, while in lower GDP level countries, factors that represent the basic needs will be more relevant to happiness. This may follow Maslow's hierarchy of needs, saying that people will first pursue the basic needs before the advanced needs in the path of pursuing their overall happiness.

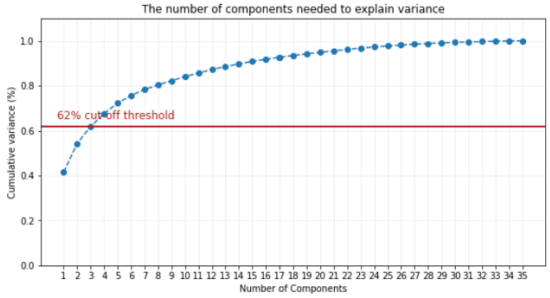
3.4 Analysis of Factors

3.4.1 Use PCA to find out which factors are the most important ones.

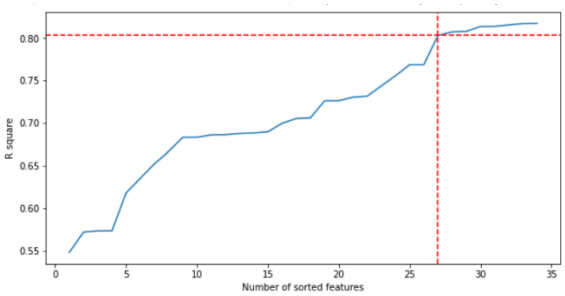


Principal Component Analysis (PCA) computes the principal components and uses them to perform a change of basis to the original data. We standardize the dataset into a data frame X with 35 external factors and calculate the eigenvalues and eigenvectors of X^TX. In the standardization process, we subtract each column of original matrix by its mean and divide the differences by its standard deviation. Then the X^TX product becomes $\frac{1}{n-1}$ times of correlation matrix of X^T with each element corresponds to a $\frac{1}{n-1}$ times correlation coefficient of a pair of factors.

The eigenvectors are the principal components which determine the directions of the new feature space, and eigenvalues determine their magnitude. Each pair of eigen vectors are perpendicular to each other. The eigenvalues explain the variance of the data about new feature axes. The scree plot from PCA indicates that the 'elbow' point where the eigenvalues level off is at PC3. The larger the eigen value, the more important the principal component is.



We graphed Cumulative variance percentage by Number of Components. PC1, PC2 and PC3 together explains 62% variance of the original data.



$$R^2 = \frac{Residual\ Sum\ of\ Square}{Total\ Sum\ of\ Squares}$$

We sort the factors by the absolute values from PC1 in descending order. The first n factors from the sorted list are trained with regression. Then we plot the R square values obtained from the regression models for 2018 against Number of sorted factors. After the first n = 27 factors, the R square reaches the largest value of 0.8033 and becomes stable. Hence, we

decided to use the first 27 sorted factors as our model input. The 27 factors are listed below:

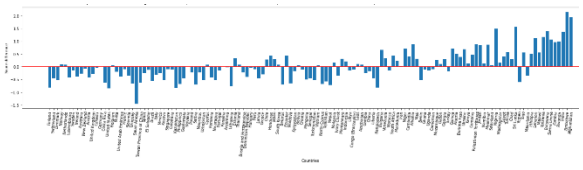
Log GDP per capita,
Life expectancy in years,
Water and Sanitation,
Women with advanced education (%),
Corruption Perceptions Index 100 = no corruption,
Child mortality rate (deaths/1,000 live births),
Health spending per capita,
Voice and accountability index (-2.5 weak; 2.5 strong),
Early marriage (% of women),
Traffic deaths (deaths/100,000),
Political rights (0=no rights; 40=full rights),
Women with no schooling,
Carbon dioxide emissions per capita,
Mobile telephone subscriptions (subscriptions/100 people),
Freedom of expression (0=no freedom; 1=full freedom),
Average annual hours worked by person engaged,
Public spending on education percent of GDP'
GINIindex (WorldBankEstimate,00-17),
Freedom of religion (0=no freedom; 4=full freedom),
International Internet bandwidth per Internet user kb/s,
Outdoor air pollution attributable deaths (deaths/100,000),
GDP per hour worked in current USD PPP adjusted,
WorldGivingIndex,
Homicide rate (deaths/100,000),
Labor force participation rate,
Homicides per 100000 people,
Positive affect.

3.4.2 Use the important factors we found to predict our own happiness scores. Compare our happiness scores prediction with the self-reported scores in the World Happiness Report.

We fit the 27 factors with multilinear regression again and applied the fitted model to 2019 testing data to predict our happiness scores. Our R square for the testing result is 0.7089. The model with 27 factors predicts the 2019 happiness score well.

Country name	Life Ladder	New Score	Rank	New Rank	Rank Difference
Finland	7.780	6.970	1.000	9.000	-8.000
Switzerland	7.694	7.252	2.000	4.000	-2.000
Denmark	7.693	7.152	3.000	5.000	-2.000
Norway	7.442	7.538	4.000	1.000	3.000
Netherlands	7.425	7.497	5.000	2.000	3.000

We can see that in our new ranking result graph, four of the first five ranking countries are still within five places among all countries, except Finland drops from 1st place to 9th place.



The plot above shows score differences between our prediction and self-reported 2019 happiness

score among countries. The scores of top two-third countries are generally under predicted, whereas the last third countries are over predicted.

4. Conclusion

In conclusion, we believe we could choose to be happy ourselves although we may not have enough in material life. Some improvements could made to raise happiness level would be, for instance:

1. Being active and make more friends, so that you will be backed by friend when things go the way making you unhappy. This will boost the social support factor for you to be able get helps from your friends.

2. Be outdoor a lot and expose to the sunlight. Latin American and Caribbean people are close to equator geographically. They have more sunlight exposures and their positive affect is so high. This contrasts to Nordic countries which has much less sunlight throughout the year.

3. Scheduling and making plans ahead, so that you will have better work-life balance.

People defines happiness differently. Our World Happiness Report considers life satisfaction as happiness, and in this definition, Nordic countries performs extremely well. Some other sources consider positive affect as happiness, like in 2014 Gallup Positive Experience Index, top 10 happiest countries in the world are all in Latin America.

5. Dataset

- The Global Economy https://www.theglobaleconomy.com.
- World Database of Happiness https://worlddatabaseofhappiness.eur.nl/
- World Giving Index https://www.cafonline.org/about-us/publications/2018-publications/caf-world-giving-index-2018
- Freedom House https://freedomhouse.org/
- World Happiness Report

https://worldhappiness.report/ed/2020/#read

References

[1] World Happiness Report 2015-2020. https://worldhappiness.report/
[2] https://www.cato.org/human-freedom-index-new
[3] Hills, P., & Argyle, M. (2002). The Oxford Happiness Questionnaire: a compact scale for the measurement of psychological well-being. Personality and Individual Differences, 33, 1073–1082.
[4] Lyubomirsky, S., & Lepper, H. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. Social Factors Research, 46, 137-155.
[5] Esteban Ortiz-Ospina (2013) - "Happiness and Life Satisfaction". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/happiness-and-life-satisfaction'
[6] J.B. Ullman, Reviewing the basics and moving forward, J. Pers. Assess. 87 (2006) 35–50.
[7] Mehdi Moeinaddini, Zohreh Asadi-Shekari, Mahdi Aghaabbasi, Ismaïl Saadi, Muhammad Zaly Shah, Mario Cools,Proposing a new score to measure personal happiness by identifying the contributing factors,Measurement,Volume 151,2020,107115,ISSN 0263-2241.
[8] Madelaine Törnquist Agosti, Åsa Bringsén & Ingemar Andersson (2017) The complexity of resources related to work-life balance and well-being – a survey among municipality employees in Sweden, The International Journal of Human Resource Management, 28:16, 2351-2374, DOI: 10.1080/09585192.2017.1340323
[9] Ejrnæs, A., & Greve, B. (2017). Your position in society matters for how happy you are. International Journal of Social Welfare, 26(3), 206-217. https://doi.org/10.1111/ijsw.12233