STUDENT PROJECT SHOWCASE

Worldwide Happiness Determinants

Iker Acha Rubalcava
Regina Alexia Blas Flores
Sergio Cuauhtémoc Blancas Vargas
Diego Alberto Zubillaga Pérez

Instituto Tecnológico de Estudios Superiores de Monterrey

Autumn 2021

1

PROJECT OVERVIEW

1 PROJECT OVERVIEW

Achieving and maintaining good health and wellness in a society is essential. Basically, wellness is an individual's willingness to feel good and function correctly, as well as in their life experiences. It is related to how the individual compares their living conditions with existing social norms and values. However, the concept of social wellness is only fully understood when several factors intertwine, it's a complex mix of social, political, psychological, environmental, economic and cultural factors that end up shaping our perception of happiness.

Happiness enables us, among other things, to better carry out those activities that will allow us to create and share value with society, so that it can develop in the best conditions possible. Due to this, and its impact on other aspects such as the quality of life, the government must ensure the well-being and happiness of the population. Therefore, for this project, we wanted to explore some databases that could shed light on the factors that affect happiness levels.

OBJECTIVE

We expect to determine the factors that affect the level of happiness of a population, and then that information can lead to the development of strategies to help achieve the following SDGs:

3 -Health and Wellbeing

Guarantee a healthy life and promote universal well-being.

16 - Peace, Justice and Strong Institutions

Access to justice for all, and building effective, accountable institutions at all levels. The objective was to determine the influence of living conditions and government involvement on the happiness of a country. In addition, the proposal was to group them according to their similarities. This could help countries make better decisions that would influence a higher level of happiness and well-being.

Our hypotheses were the following:

Freedom influences happiness positively

Life expectancy influences happiness positively

Corruption influences happiness negatively

DATABASES USED

World Happiness Report up to 2020: This file contains happiness scores of 153 countries, as well as factors used to calculate the 2015-2020 score. That happiness score is a national average of responses to the primary life assessment questions formulated in the Gallup World Poll (GWP), which uses the Cantril scale.

Rule of Law Index: This database includes the scores of the factors evaluated by the Rule of Law index (limits to governmental power, absence of corruption, open government, fundamental rights, order and security, regulatory compliance,civil justice, and criminal justice) of each country from 2012 to 2013.

The World Justice Project Rule of Law Index 2020. Rule of Law Index 2020 [Database].
Retrieved from https://worldjusticeproject.org/our-work/research-and-data/wjp-rule-law-index-2020

World Happiness Report. World Happiness Report 2015-2020 [Database].
Retrieved from https://worldhappiness.report/archive/

2

PROCESS & USE OF DSS

2 DATA CLEANING MODEL

The Rule of Law databases contained the countries in columns, so they had to be transposed with the rows. In these rows, we had each Factor broken down into sub-factors that were taken into account to determine it (i.e. Factor 1 was made up of 1.1, 1.2...), so we only consider that main Factor. The region and the Income Group were also not required, so we decided to ignore them.

The following factors were used and shortened to 'Fs' to avoid misunderstandings: F1 - Constraints of government powers, F2 - Absence of corruption, F3 - Open government, F4 - Fundamental rights, F5 - Order and security, F6 - Regulatory enforcement, F7 - Civil justice, F8 - Criminal justice. This procedure was applied to the 5 databases of the Rule of Law.

In the Happiness Index databases, the categories of countries (Country), Happiness Score and the 6 variables that determined it were maintained: C1 - Economy (GDP per Capita), C2 - Family, C3 - Health (Life Expectancy), C4 - Freedom, C5 - Trust (Government Corruption) and C6 - Generosity. They were also abbreviated with a 'C' to avoid misunderstandings. The other variables in the datasets were ignored.

Country	Happiness Score	C1-2015	C2-2015	C3-2015	C4-2015	C5-2015	C6-2015	F1_2015	F2_2015
string	double	double	double	double	double	double	double	double	double
Country	Decimal	Decimal	Decimal	Decimal	Decimal	Decimal	Decimal	Decimal	Decimal
Denmark	7.527	1.32548	1.36058	0.87464	0.64938	0.48357	0.34139	0.9181954932534434	0.9560304365721386
Norway	7.522	1.459	1.33095	0.88521	0.66973	0.36503	0.34699	0.8756678700116582	0.931957844028702
Canada	7.427	1.32629	1.32261	0.90563	0.63297	0.32957	0.45811	0.7788059983214639	0.8144293318076585
Finland	7.406	1.29025	1.31826	0.88911	0.64169	0.41372	0.23351	0.8809414467341535	0.9011585348798758
Netherlands	7.378	1.32944	1.28017	0.89284	0.61576	0.31814	0.4761	0.8724775234890959	0.8885504212651943
Sweden	7.364	1.33171	1.28907	0.91087	0.6598	0.43844	0.36262	0.8754877723972708	0.905332613180136
New Zealand	7.286	1.25018	1.31967	0.90837	0.63938	0.42922	0.47501	0.8464481802440815	0.898947542823636
Australia	7.284	1.33358	1.30923	0.93156	0.65124	0.35637	0.43562	0.8300835435031041	0.8445498892106584
Costa Rica	7.226	0.95578	1.23788	0.86027	0.63376	0.10583	0.25497	0.7831476301341503	0.6808352199427374
Austria	7.2	1.33723	1.29704	0.89042	0.62433	0.18676	0.33088	0.8485560181379322	0.8298368276549978
Mexico	7.187	1.02054	0.91451	0.81444	0.48181	0.21312	0.14074	0.5146215461972731	0.3337439713036612 4
United States	7.119	1.39451	1.24711	0.86179	0.54604	0.1589	0.40105	0.7615390765426393	0.7546570153554917
Brazil	6.983	0.98124	1.23287	0.69702	0.49049	0.17521	0.14574	0.6095144783952386	0.4600986608856613 7
Belgium	6.937	1.30782	1.28566	0.89667	0.5845	0.2254	0.2225	0.8087203469377177	0.8074321424750823

Example of a Cleaned Database

The databases were initially limited by the number of countries. There are around 200 countries, depending on the source since the recognition of some nations is not universal (i.e. Israel, Taiwan). The second limitation of the information is that several countries do not have the financial means to carry out such investigations. The third limitation is that several countries decide not to report this data to beautify the country's situation before the international community. And finally, there may be a confusion between the items that have to be presented and there may be inconsistencies or errors in how the information was captured.

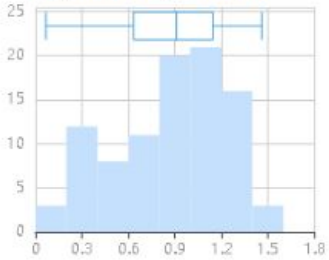
It is for this reason that using the Dataiku tool we carry out a deep analysis of the databases, thus discovering that there is a great relationship and similarity between the information in the happiness database of 2017 and that of 2018. Therefore, it was decided to do a joint analysis between those two.

STATISTICAL ANALYSIS

For the descriptive analysis, the different variables were analyzed through the years (2015-2020), obtaining certain statistical measures that describe the behavior of the variables. In this case, the mean, median, standard deviation, variance, kurtosis, asymmetry coefficient, and range were obtained, together with the corresponding histograms and some adjustment tests to determine the probability distribution.

To determine the normality of each variable, the shape of the histogram was first observed, which would give us clues about the possible probability distribution, and then we proceeded to carry out a normality test. In this case, the Kolmogorov-Smirnov test was used with a confidence level of 95%, which means that the p-value has to be greater than 0.05 to say that there is no statistical evidence that the variable is not normal. If it is less than 0.05, the histogram will be re-examined to propose possible probability distributions, and the corresponding tests will be carried out.

Fig.13 GDP per capita



Average: 0.872047128
Median: 0.910295
Standard deviation: 0.363064863
Sample variance: 0.131816095
Kurtosis: -0.663360571
Asymmetry coefficient: -0.440601534
Range: 1.50582

Normality test by Kolmogorov-Smirnov
p-value: 0.5391431795

Example of the Statistical Analysis

Through the pertinent tests, it was possible to determine the distribution of all the variables studied, and with their results, it could be ensured that most of the variables were normally distributed. However, we also found other distributions such as Beta, Triangular, Exponential, and Laplace, in which some variables fit for certain years.

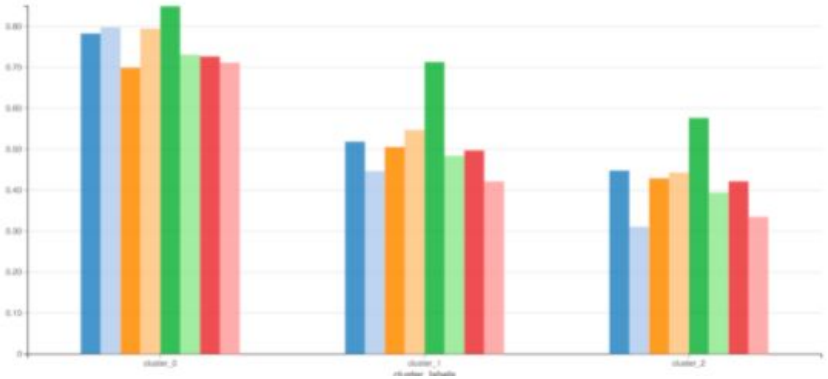
The cleaning of the datasets meant that in the end we had less data, but thanks to this we were able to carry out these analyzes in a congruent way and using the same variables. The determination of these statistical indicators allowed us to have a better understanding of the information we would introduce to the models, and consequently, to have a better understanding of the results.

CLASSIFICATION MODEL

For the classification model, tests were carried out with 4 different models to choose the best one. These were: Agglomerative Clustering (one of 3 groups and another of 5 groups), Gaussian Mixture, and K Means of 5. The best of the 4 was Agglomerative Clustering of 3. Given that there were few variables of numerical type, the R2 was 0.269, which for our model was effective. So as not to assume that the groups would be maintained by year, the same model choice test was carried out for each year (2015-2020). In all cases, it turned out that Agglomerative Clustering was the most appropriate.

To facilitate interpretation, the results for each year were graphed in three different ways: two histograms, one for indicators of the rule of law and another for indicators of happiness, and world maps with the clusters highlighted.

The maps were made with the Geocoder plugin. First joining each country with its latitude and longitude, then extracting the GeoPoint address of the country with that data, and finally graphing it using its GeoPoint.



Example of the rule of law indicators by group of countries in 2015

Fig.92 Cluster 0 of 2015

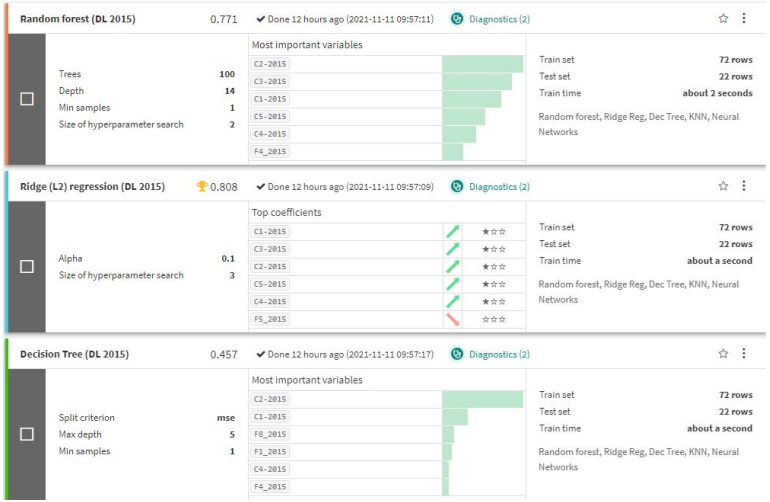
It mainly encompasses developed countries with high confidence in the government, social programs, and high GDP.

They are mostly found in western and northern Europe. In addition, many were the headquarters of colonialist states.

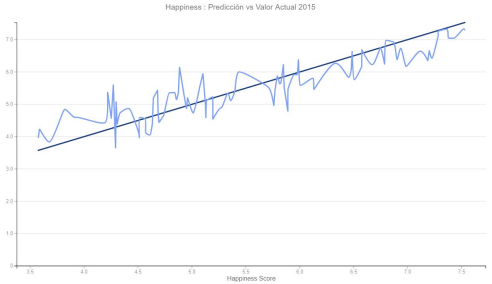
Example of map with clusters highlighted

PREDICTIVE MODEL

To choose the prediction model, tests were first carried out with 3 different models in Deep Learning: Ridge Regression, Decision Trees, and Random Forests. The chosen model was Ridge Regression because it had the largest R2 value.

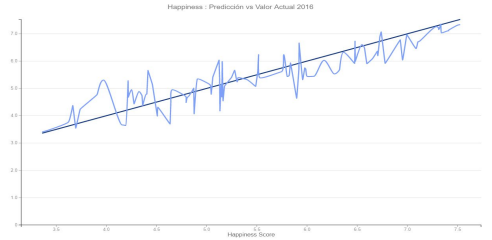


Comparison of Prediction models



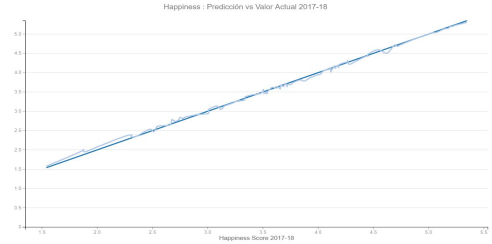
2015
Prediction vs Present Value

As can be seen, we have too many fluctuations for predicting happiness, not to mention abrupt changes, although this may be an increase in happiness in those countries, it can also mean a rapid fall.



2016
Prediction vs Present Value

In this graph the fluctuations and changes are even more abrupt, we can say that this prediction model is not the most appropriate. Also the mean is lower than the previous one, which indicates a decrease of happiness.



2017-18
Prediction vs Present Value

We can observe a somewhat linear trend, with few fluctuations and with a marked positive rise for the increase in happiness in the countries analyzed.

2020
Prediction vs Present Value

In this last prediction we notice a smaller distance between each peak, which indicates, with respect to the previous year, happiness will have an even greater instability.

3

OUTCOMES & CONCLUSION

3 OUTCOMES & CONCLUSION

According to the results obtained in the simulations, we were able to face the hypothesis and conclude that:

Freedom positively influences happiness

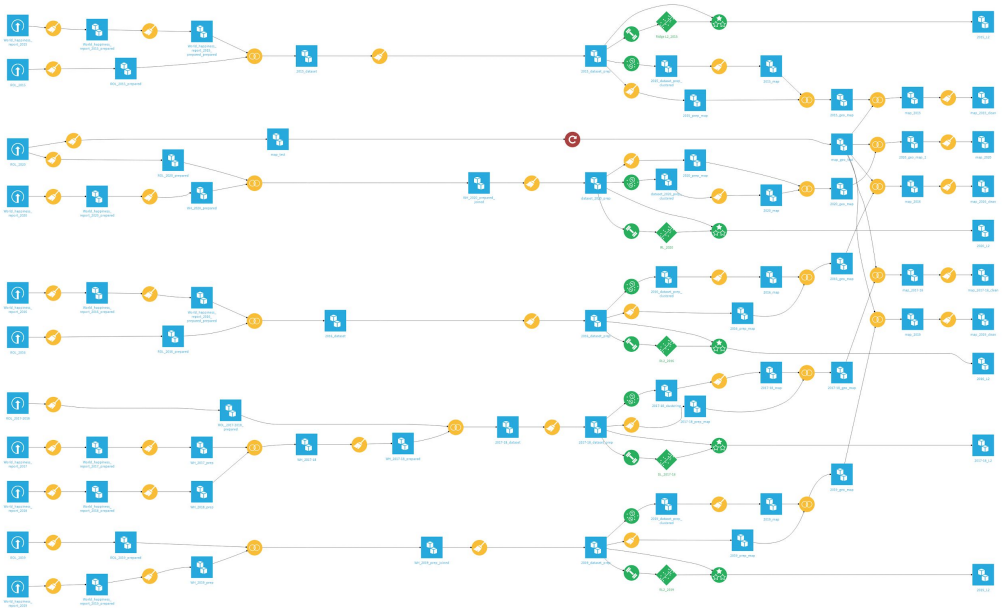
It claims to be true because we found a strong relationship between these two variables. However, there are several cases where countries with less freedom are still happy.

Life expectancy positively influences happiness

It was also found to be true, there is a close relationship between the two. Notably, there were only a few that did not follow this pattern, especially in Africa.

Corruption negatively influences happiness

This hypothesis was refuted, since there are several countries, like Mexico, where corruption is high but still there is a high degree of happiness.



View of the Complete Model

The analysis, no matter how exhaustive, has a significant bias. Half of the countries in the world were not used for the study, since their data was not complete for the analysis. There is a serious lack of information gathering on their behalf. Also, when the information is shared, it is usually not truthful or has been modified in some way by the government, criminal groups, or private companies. This leads to showing your country with a less critical eye.

Assuming that we had data from all the countries, we would still have to increase the number of variables to have a scenario that better reflects reality. The number of countries is limited, with just a little over 200, and in most models, more data is needed for better analysis.

TEAM STATEMENT

“A better world allows us to live out a better life”

Anonymous

Our project should be selected as the overall winner because of the importance the world places on happiness. As stated before, happiness is more than just feeling good, it enables us to better perform in those activities in which we create value for the society, being an important factor for development and quality of life. For that reason, governments should not undermine the potential of a happy population, and take advantage of technology to make it happier. Therefore, we believe our project lays an important step to understand how governments can increase happiness levels, starting from data analysis and then lead the insights to the development of relevant strategies and policies for a better society.

The benefits of performing this type of analysis is that these decisions have a data background, which ensures that strategies leverage upon key factors to address the objectives, resulting in real and tangible changes. For that reason, we consider important to promote this kind of approach, that can be extrapolated to other problem areas such as health and safety.

Research from Harvard University in 2017 revealed that people who are happy are in better health. This study has been going on for several years and has shown that happy people sleep better. Happiness helps slow mental decline, controls stress, and helps with memory capacity. When many people talk about the importance of happiness, they immediately associate it with money, fame, and success. However, what this study revealed is that happiness is provided by good interpersonal relationships.

This finding showed that good personal relationships, that is, creating emotional ties, has a great impact on our health, bringing great benefits on it. The study affirms that people who feel lonely are at higher risk of developing health problems. This research is endorsed by the American Association of Psychologists, where they stated that loneliness is a public health risk. Based on the Harvard study and what we found in ours, we could say that happiness makes us more optimistic, less stressed and pessimistic. It makes us enjoy better physical and mental health, in addition to enjoying good personal relationships, have a longer live and perform better in our daily activities.

Mineo, L. Good genes are nice, but joy is better. The Harvard Gazette [2017]. Retrieved from https://news.harvard.edu/gazette/story/2017/04/over-nearly-80-years-harvard-study-has-been-showing-how-to-live-a-healthy-and-happy-life/
Silwa, J. So Lonely I Could Die. American Psychological Association [2017]. Retrieved from https://www.apa.org/news/press/releases/2017/08/lonely-die

Thank you! - ¡Gracias!