World Happiness Report

David D’Onofrio, Liza Kottler, Annika Fagerstrom and Rachel Konshok

Fordham University

# Abstract:

The World Happiness Report is a landmark survey of 155 world countries that has been presented every year since 2012 to the United Nations. The data set from this survey can be used to investigate the factors that go into happiness at national and international levels, and study trends and how they change through the years. Studying this data set can give insights into what factors most dramatically impact happiness as well as what factors actually contribute very little to national happiness. The goal of this study can be split into three main aims: by using the data given, create a mechanism to predict future happiness scores and the direction in which they will change; find the strongest attributes affecting the happiness score; finally, to understand how GDP per capita and the happiness score are related.

# 

# 

# 

# Introduction:

The World Happiness Report (WHR) is a report presented to the United Nations every year in March that attempts to measure in a quantitative way the happiness of the countries of the world and study the factors that contribute to a country’s happiness. This survey is conducted by Gallup every year using a Cantril ladder method to get numerical data for what often seem to be qualitative questions. This report and associated data set have been prepared every year since 2012 and provide not only a look into any countries’ happiness but also at how happiness and the factors that affect happiness change over time both at national and international levels.[[1]](#footnote-0) Since the survey has been repeated every year with the same methods and data collected, it is an excellent way to study change over time for many different factors.

The world happiness report data generates a final happiness score for all 155 countries in the study from a series of contributing factors. These factors are GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, family life, and perceptions of corruption. We didn’t use the attributes of family life or social support, as unfortunately data on them was not available for both 2018 and 2019. Population per country was added into the data set in order to examine the relationship, if any between population and overall happiness.

For every country’s score in the data set, both attribute scores as well as final happiness score were recorded.. Within each country, the UN team surveys 2,000 to 3,000 residents of that respective country.[[2]](#footnote-1) Although this may seem like a small amount of data points to generate a score for an entire country, especially large countries, almost every country has a 95% confidence interval.[[3]](#footnote-2) Most participants are recruited through random phone calls, Gallup uses either randomly generated phone numbers for a country or a representative sample of registered phone numbers.[[4]](#footnote-3) There are some countries where less than 80% of the population has phone numbers, and in such a situation, Gallup will instead visit in person to ask the same questions.[[5]](#footnote-4)

The Cantril Ladder is a method of measuring personal happiness in life through a series of questions and it gives numerical values as results as opposed to qualitative information. It involves asking participants to imagine a ladder with ten ladder rungs from zero to ten, with the worst possible and most miserable life scenario at zero and a perfectly happy world at ten. For each question, a participant is asked to identify on what rung they would place their own life in relation to a specific question. This method gives numerical values as the answer to every question and ensures that the data can be averaged and worked with easily as a result.[[6]](#footnote-5)

One of the attributes of the data set is a numerical measurement of the perception of corruption for a given country. This attribute may not initially seem to be the easiest to quantify. Gallup solves the issue of qualifying corruption by asking every participant two binary questions relating to corruption. The questions are “Is corruption widespread throughout the government or not?” and “Is corruption widespread within businesses or not?”.[[7]](#footnote-6) Since these are binary questions and not Cantril ladder formatted questions, the results from all participants in a given country are averaged and then normalized to fit with the rest of the attributes being measured on a zero to ten scale.[[8]](#footnote-7)

One part of the WHR data set is the concept of ‘Dystopia’, a fictional country invented for the data set. This country is the country with the least happy people on average of all countries in the world. It’s attribute scores are based on the averages of the lowest countries to give it the lowest score in every attribute as well as the total score. This means that Dystopia will act as the baseline score for the set and therefore every single country can be favorably compared to Dystopia.

The world happiness data set comprises a final happiness score, as well as several attributes. These attributes are GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption as well as the added column of population, as well as a rank of the countries from least to most happy for 155 countries. These are the attributes that comprise the data set that was pulled from the Kaggle from the 2018 and 2019 years of the survey.

By studying the WHR data set over two sample years, 2018 and 2019, not only can the attributes contributing to happiness be analyzed by impact but also by change over time. This change over time allows the study into predictions of future changes of the happiness score both overall as well as for an individual country.

For the purposes of this research, three research goals were identified. Goal 1: Predict if happiness score will go up or down. Initial impressions were that this would be likely difficult to prove universally, but interesting to study. Goal 2: Find strongest features contributing to happiness. This goal was assumed to be fairly straight forward to examine through feature selection methods. Goal 3: Examine the relationship between GDP and Happiness. This final goal was created via an assumption that GDP would prove to be one of the most important features identified in feature selection, and the link between countries wealth and happiness is frequently discussed both nationally and internationally, but often without much data.

# Results

The initial pre-treatment for the data involved a merge with some population data, and then identifying outliers, done via a left merge to keep the initial WHR data set complete and add only the requested data from the secondary data set. The population data set included came from a different United Nations report, the World Population Prospects report. This report contains population data starting from 1950 and is up to date to 2020. After 2020, the data set makes projections about the world populations (as well as other attributes that were not included in the merged dataset) until 2100.[[9]](#footnote-8)

This merge of the World Happiness Report data with the World Population Prospects data resulted in around 30 new missing values that needed to be filled in. The original WHR data only had a single missing value that was simple to fill, these new missing values required more work in order to plug the missing values. These values were filled post-binning with the mean of the population for the bin that country was placed into. This meant that the missing values were filled by the mean of countries that were very similar, and should give a filled value result with strong confidence.

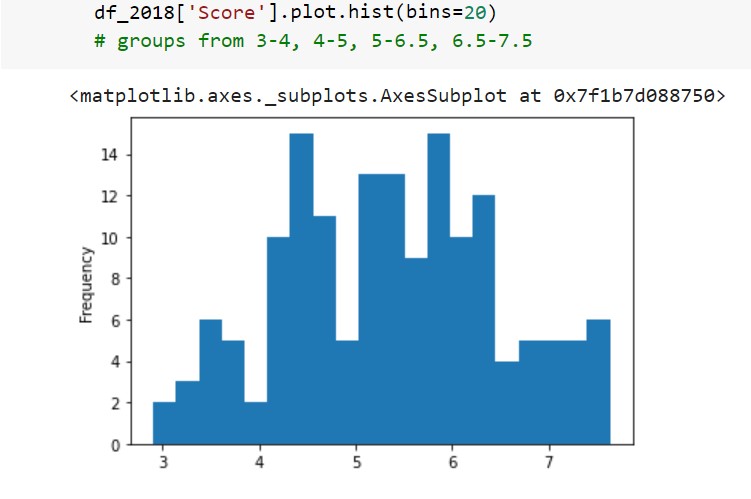
Binning was done by width as opposed to height around the values for overall happiness score. This keeps bin groups filled with data points with similar values and therefore hopefully similar behavior. The choice was made to bin by width over by height to keep some bins from just holding outlier data or a mix of data over a very wide range of happiness scores. Both of which would make it harder for later modeling methods to work successfully.

Fig 1: The data visualized into 20 bins by width

Since for this project both the 2018 and 2019 data sets were in use, the data sets were merged to create a single large data set with two columns for every attribute - a 2018 and 2019 set for every attribute. One additional column was created during this process, a ΔScore column which was calculated by the change of the happiness score between the 2018 and 2019 scores. A positive change in the ΔScore column for a country indicated that between 2018 and 2019 the country’s happiness score decreased. And, if the ΔScore value was negative, that would indicate that a country’s happiness score went up between years.

Applying the association rules analysis algorithm revealed some interesting relationships for the data within their respective bins. The association rules algorithm finds connections between items within a data set that frequently occur together. For this particular data set, the association algorithm will identify strong relationships between different attributes but not necessarily correlation. Ie. the relationship between GDP and health outcomes can be shown to be linked as a result of this algorithm, but not conclusively proven to be connected in a form like ‘high GDP guarantees high health outcomes for a county.’

Table

Description automatically generated

Fig 2: The association rules table generated with a support of 10%

The first step for feature selection is Naive Bayes algorithm. Naive Bayes is a classifier algorithm that can be scaled easily with different numbers of features selected to find the highest accuracy. For the WHR data set, the algorithm was tested for multiple different numbers of features selected. Ultimately, the algorithm tested with selecting 10 features was found to have the highest accuracy of all variations tested, with an accuracy of 69%.

The next step of the data analysis was to perform feature selection algorithms. This first of which is the chi squared algorithm which takes the features of a given data set and determines which ones have the greatest impact on the final score category. This algorithm does so through testing hypotheses to determine expected outcomes for the data and discover the attributes with the strongest impacts. Using the number of features as selected by the Naive Bayes algorithm, the Chi2 algorithm selected the ten most impactful features for the data set.

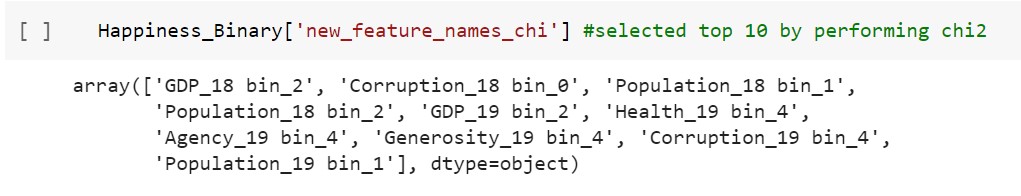


Fig 3: Top 10 features as selected by Chi2 algorithm

Once the top 10 features are established, we split the data into a training and testing set (70% training, 30% testing). All models are tested using the same training and testing sets for consistency. The next algorithm creates a decision tree (criterion = ‘entropy’) starting with a root node and splitting the set around a variable. From there, it builds out nodes by repeating the processes and splitting data on the next variable. This modeling method can predict where new data points will end up on the attribute the model is trying to predict by following through the tree for the given attributes. The decision tree was calculated using seven different combinations of maximum depth and minimum samples leaf values. The tree is Fig. 4 has the highest score and smallest standard deviation with a max\_depth = 3 and min\_samples\_leaf = 5 (score = 0.68085, std = 0.62). Based on this decision tree, the top five features in order are:

#1 = Population\_18 bin\_1 (5083.75, 11328.5], #2 = Corruption\_18 bin\_0 (-0.001, 0.051], #3 = Population\_18 bin\_2(11328.5, 33396.25], #4 = GDP\_18 bin\_2 (0.95, 1.198], #5 = Population\_19 bin\_1 (5213.5, 11531.0].

Diagram

Description automatically generated

Fig 4: The Best Decision Tree (max\_depth = 3, min\_samples\_leaf = 5)

The next models trained are using the Naive Bayes prediction model. Since the data was all categorical, we trained the Bernoulli Naive Bayes model and the Multinomial Naive Bayes model. After training, the models were tested and predicted with an accuracy of 0.67 (+/- 0.04) and 0.69 (+/- 0.07), respectively. Therefore, we concluded the Multinomial Naive Bayes model is better.

The final models trained are using the ensemble classification method. We trained both a random forest classification and a bagging classification model. After training, the models were tested and predicted with an accuracy of 0.63 (+/- 0.10) and 0.66 (+/- 0.15), respectively. Therefore, we concluded the bagging classification model is better.

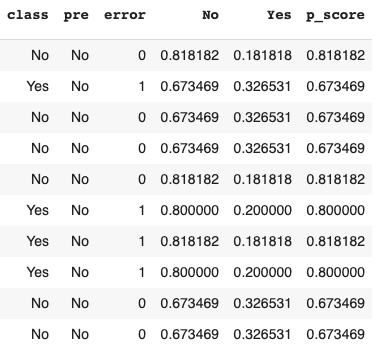
After we had selected the ideal model for each prediction method, we created two new columns for each model: the class prediction and error rate, whether the model is right(0) or wrong(1). These columns can be seen in Fig 5,6,7 below.

Fig 5 (to left): Decision tree prediction and error

rate.

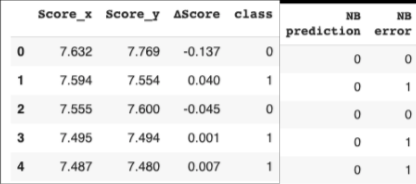


Fig 6 (to right): Multinomial Naive Bayes prediction and error rate.

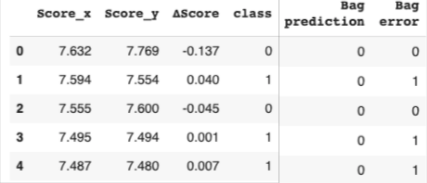


Fig 7 (to left): Bagging ensemble classification prediction and error rate.

To compare the performance of all three models, the mean of the error rate columns were calculated. The decision tree error rate has a mean = 0.3191 (std dev = 0.4712). The multinomial naive bayes error rate has a mean = 0.3077 (std dev = 0.4630). The bagging ensemble classification error rate has a mean = 0.2949 (std dev = 0.4575). With the lowest error rate mean and standard deviation, the bagging ensemble classification error rate comes out as the best prediction model.

# Results

So what are the conclusions we can draw from these outcomes? Thinking about this in relation to our original goals, with the first goal being to use existing data from 2018 and 2019 to predict the happiness score and whether it would go up or down in the future. Looking at the data it was clear predicting future happiness scores would not be possible as they are the sum of all other categories, therefore to predict them one would need to accurately predict each category and the data did not go back far enough temporally to accurately predict that. Therefore, the goal was changed to instead focus on predicting the directional change of the happiness score. Through the feature selection performed through the Chi squared test, and confirmed with the Naive Bayes test, the top ten most important features were selected that actually improved the overall accuracy from the original dataset (from .60 to .69). From this we used several different models to predict the change and found that bagging returned the lowest standard error. Therefore using this code we could see for each country there future probable direction in happiness score, so this goal was achieved.

The second aim of this experiment was to find the strongest attributes affecting the happiness score. This was done through feature selection. Again using the chi square test paired with the naive bayes to test for accuracy we were able to determine that the strongest attributes were: GDP from 2018 bin 2, Perceptions of Corruption from 2018 bin 0, Population from 2018 bin 1, Population from 2018 bin 2, GDP from 2019 bin 2, Healthy Life Expectancy from 2019 bin 4, Agency from 2019 bin 4, Perceptions of Corruption 2019 from bin 4, and Population from 2019 bin 1 respectively. Thus our second aim was achieved.

The final aim of this study was to understand how GDP per capita and the happiness score were related. We initially assumed that we would see a high correlation between GDP per capita and a strong happiness score. We found this to be true from two of our results. First with feature selection two of our top ten features are GDP for 2018 and 2019 indicating that GDP highly affects the final happiness score. Secondly, when performing association analysis GDP bins appeared more frequently than any other attribute indication that high GDP is an antecedent of high healthy life expectancy and consequently high happiness scores. So with that our final aim of the study was achieved.

A particular area that could perhaps be improved in future experiments include handling of interesting outlier cases, we had two of these cases both with Perceptions of Corruption and Population. In both these cases the number of upper bound outliers was 14. With Population we merely corrected this with a new upper bound but we decided not to do this with Perceptions of Corruption. The reasoning for this was that we found Perceptions of Corruption a truly poignant case in which we didn’t want to normalize and country’s data for that attribute. If someone marked Perceptions of Corruption abnormally high we wanted that reflected in the final calculations as that should lead to a significant change in the Happiness Score. Whereas with Population, while it is an interesting attribute, it is not decided upon by the takers of the survey. Overall the takers had no say over Population as that is indisputable, but Perceptions of Corruption is an important characteristic decided upon by the survey takers. Overall this could be changed in future experiments, but mostly it was just an interesting case that we believed required special thought.

**Conclusion:**

*Goal 1 : Predict if happiness score will go up or down.* Ultimately, studying the data proved that for any given country it is possible to predict a change of some kind. Through several prediction models (Decision Tree, Naive Bayes, Ensemble - Bagging, and Random Forest) the ability to predict change was confirmed.

*Goal 2 : Find strongest features contributing to happiness.* Through feature selection and entropy feature importance measurements the most impactful features on the final score. These features were found to be GDP, population, and the perception of corruption. The most interesting inclusion on this list is the presence of population. This is the additional attribute added into the data set at the beginning of pre-processing in order to investigate the relationship between population and a country’s happiness.

*Goal 3 : Examine the relationship between GDP and Happiness.* In both the feature selection and association rules modeling, GDP showed up as a large factor contributing to happiness score. In fact, in association rules modeling, different GDP bins showed up as ten of the antecedents out of 13 generated rules. Interestingly, the consequences of the majority of rules with the GDP as an antecedent, was the attribute ‘Health’. So more than finding a strong correlation between GDP and overall happiness of a country, the data also showed a very strong relationship between high GDP and higher health outcomes for a country.

# References:

“FAQ.” *World Happiness Report*, worldhappiness.report/faq/.

Gallup. “Understanding How Gallup Uses the Cantril Scale.” *Gallup.com*, Gallup, 8 Apr. 2021, news.gallup.com/poll/122453/understanding-gallup-uses-cantril-scale.aspx.

John F. Helliwell Vancouver School of Economics at the University of British Columbia, and Vancouver School of Economics at the University of British Columbia. *Changing World Happiness*, 20 Mar. 2019, worldhappiness.report/ed/2019/changing-world-happiness/.

Sustainable Development Solutions Network. “World Happiness Report.” *Kaggle*, 27 Nov. 2019, www.kaggle.com/unsdsn/world-happiness.

1. Sustainable Development Solutions Network. “World Happiness Report.” *Kaggle*, 27 Nov. 2019, www.kaggle.com/unsdsn/world-happiness. [↑](#footnote-ref-0)
2. “FAQ.” *World Happiness Report*, worldhappiness.report/faq/. [↑](#footnote-ref-1)
3. “FAQ.” *World Happiness Report*, worldhappiness.report/faq/. [↑](#footnote-ref-2)
4. ibid [↑](#footnote-ref-3)
5. ibid [↑](#footnote-ref-4)
6. Gallup. “Understanding How Gallup Uses the Cantril Scale.” *Gallup.com*, Gallup, 8 Apr. 2021, news.gallup.com/poll/122453/understanding-gallup-uses-cantril-scale.aspx. [↑](#footnote-ref-5)
7. John F. Helliwell Vancouver School of Economics at the University of British Columbia, and Vancouver School of Economics at the University of British Columbia. Changing World Happiness, 20 Mar. 2019, worldhappiness.report/ed/2019/changing-world-happiness/. [↑](#footnote-ref-6)
8. John F. Helliwell Vancouver School of Economics at the University of British Columbia, and Vancouver School of Economics at the University of British Columbia. Changing World Happiness, 20 Mar. 2019, worldhappiness.report/ed/2019/changing-world-happiness/. [↑](#footnote-ref-7)
9. WPP info [↑](#footnote-ref-8)