

Taylor Imhof

Bellevue University / DSC 520

Final Project Step 1

11/11/2021

INTRODUCTION

Happiness is such a subjective state, and it is one that everyone seeks to attain. It is very complex, and many factors contribute to one's perception of happiness. For example, some people find fulfillment in their occupations to be a large contributor to their happiness, while others might attain happiness in the service of others. I have always been curious about what specific factors contribute the most to overall happiness. Since happiness cannot be achieved with a "one-size-fits-all" approach, I thought it would be interesting to see if I could utilize my newfound data analytics skills to identify the most effective happiness factors.

When forming a potential data science question around this idea, I initially thought researching what makes people happy would be a good start. However, I quickly found that this would be much too broad of a question. So, instead, I decided to frame the research problem as to answering the question, "how can people increase their likelihood of being happy." I felt that if I could come up with a few insights based on my findings, I could prescribe to my audience which circumstances are more likely to result in increased levels of happiness.

People should be interested in this project because if we can use the results to increase the global population's mean happiness level, we will have achieved something spectacular. Furthermore, since I will be utilizing data sets that measure different aspects of human nature

that correspond with happiness, this should be considered a data science problem. Also, I anticipate reporting my results as probabilities or odds of being happy given certain conditions.

RESEARCH QUESTIONS

1. What is happiness?

As discussed above, happiness is a subjective state. As such, I feel the best way to measure happiness would be to report survey findings. I found a great list of datasets containing the results from the World Happiness Report (outlined in the datasets section) that provides an overall score for a country that I felt would be useful.

2. Which countries are the happiest?

I felt that this would be one of the easiest questions to answer, given the primary data set I plan on using for this project. The World Happiness Report seems to have been primarily used to depict the happiest countries and rank them accordingly. For instance, the top three happiest countries in 2018 were Finland, Norway, and Denmark (Helliwell, Layard, & Sachs, 2020). While not a perfect example of how I would expect the findings of this project to be used, but one could say by moving to either of those three countries, you have a higher probability of being happy.

3. How much of a role does physical/mental health play in feelings of happiness?

During my time in the Air Force, we were beaten over the head with the concept of the “four pillars of wellbeing.” While it certainly got repetitive doing biannual resiliency training, I do feel they have great merit. The four pillars are mental, physical, social, and spiritual (Biese, 2017). I wound up finding a pretty decent data set concerning wellbeing and lifestyle factors. I thought it

would be interesting to include this in the project as it is more granular in that it reports survey results from individuals, rather than results by country like with the World Happiness Report.

4. Are countries with higher rates of employment happier than those with lower?

While the GDP is a good metric for measuring economic prosperity, it does not provide as much detail to a country's individual economic circumstances. I felt it would be useful also to provide the employment rates of each of the country's outlined in the World Happiness Report

5. Do countries with stronger social programs have happier populations?

While I am certainly not an expert on the matter, I feel that countries with stronger governmental support programs are likely to have a happier populace. To measure this, I decided it would be useful to investigate how much countries allot to their social spending budgets.

APPROACH

To answer my proposed question of how one would increase their chances of being happy, I will analyze datasets to find certain trends that might increase the likelihood of feeling happy. By using the questions outlined above as a foundation, I feel that I should be able to break down the results into prescribable recommendations that I would present to my audience. Then, through data exploration, transformation, and summarization, I expect to make general statements about which factors and circumstances increase the probability of being happy.

DATASETS

1. World Happiness Report

The data contained in these datasets are derived from life evaluation questions gathered in the Gallup World Poll (GWP). The first report was released in 2012 and has since gained global respect as indicators to assist policy-making decisions. Apart from a happiness score derived by participants responding using the Cantril Scale (CS) that measures general life satisfaction, the dataset includes GDP, life expectancy, freedom, and social support from 153 countries. There are 20 columns included, and all data appears to be clean and valid, though further examination will be conducted.

Original Source – World Happiness Report (Helliwell, Layard, & Sachs, 2020)

2. Lifestyle and Wellbeing Data

This data set contains responses from nearly 16,000 survey participants concerning work-life balance. Aside from general metrics such as age and gender, this data set includes answers to general life questions like the number of close friends, typical hours of sleep, and hours working on a passion. The answers to the survey (which vary in response format) are then used to calculate an overall work-life balance rating. Therefore, I felt that this dataset would potentially provide great insight into what people can do individually to increase their feelings of happiness. There are approximately 16,000 rows of data with 24 columns, all of which appear to be already clean, though further examination will be conducted.

Original Source – I believe the dataset is maintained by the Kaggle author (Yvon Dalat). I could not find how he goes about pulling in results from the survey data, but he does seem to be actively maintaining the data repo on Kaggle. (Dalat, 2021)

3. Human Development Report 2015

For some additional insights into happiness factors at the country level, I found an interesting dataset that provides additional columns that might be useful. The primary metric outlined in this dataset is the human development index (HDI), a summary measure of achievements in several key dimensions, including life expectancy, years of education, and gross national income. The dataset also included a GDP column, which should line up with the GDP column in the World Happiness Report dataset. There appear to be around 190 rows of data and only a few missing values, which will have to be handled.

Original Data Source – United Nations Development Programme (United Nations, 2021)

REQUIRED PACKAGES

A common trend that I have seen in many data science guides and tutorials that utilize R as the primary programming language is the use of the group of packages contained in the tidyverse, which includes ggplot2, dplyr, tidyr, readr, purrr, tibble, stringr, and forcats. Ggplot2 will be used to create the different types of graphs and visualizations of the project's data. Dplyr will be used for data manipulation and getting used to the structure of the data. If looping through the data is required, I will use functions from the purrr package. After briefly reviewing the columns in my data, there are a few that contain strings. If I feel that they are worth examining and string manipulation is warranted, I will call upon functions from the stringr package.

PLOTS AND TABLES

When first exploring my datasets, the most common type of plot that I will likely be creating is simple histograms to see how each variable is distributed. If I find that some variables

appear to be related, I will create scatterplots to see if I can see any linear trends between the data. Summary statistic tables will help me discover useful observations about how each variable distributes in the population. If I conduct any correlation analysis, then the coefficient tables will be useful to highlight significant relations or trends.

QUESTIONS FOR FUTURE STEPS

One process that I predict having to complete for this project is combining columns from two or more data sets for comparison. I believe this was covered briefly early in the course, but I will certainly need review to implement it as required for this project properly. In addition, there are quite a few columns that I will need to refer to the source material and specifications in order to understand what the values equate to fully. Finally, if necessary, I might have to recode some of the values in the columns to perform the analysis required to answer my research questions.

As I go about working through this final project, there will likely be additional questions or concerns that will require me to reevaluate the research questions. In that scenario, I will be sure to note any changes to the core question or objective in future project steps.

References

- Biese, T. (2017, November 20). Building resiliency with right tools. *Military News*. Retrieved from https://www.militarynews.com/peninsula-warrior/news/air_force_news/building-resiliency-with-right-tools/article_f462c3fa-7797-541b-bd89-8e46f04ae10f.html
- Dalat, Y. (2021, March 14). Lifestyle_and_Wellbeing_Data. *Kaggle*. Retrieved from https://www.kaggle.com/ydalat/lifestyle-and-wellbeing-data?select=Wellbeing_and_lifestyle_data_Kaggle.csv
- Helliwell, J., Layard, R., & Sachs, J. (2020). World Happiness Report 2020. *New York: Sustainable Development Solutions Network*. Retrieved from <https://worldhappiness.report/ed/2020/>
- United Nations. (2021). Human Development Reports. *United Nations Development Programme*. Retrieved from <http://hdr.undp.org/en>