

Matthew Ko

## What Contributes to Happiness the Most?

Happiness is a concept that everyone ultimately strives for, but there is no one checkbox such as wealth, family with children, or healthy that someone can check as true and always consider themselves happy. Being in a state of happiness is universally recognizable to most individuals, but sometimes we do not even know exactly why we are happy or what we need to do to become happy. Happiness is usually difficult to measure directly on its own as many factors weigh into feeling happy. My goal with this project is to give a more concrete definition of what makes people happy in terms of more quantifiable terms.

The dataset I will be exploring is from the European Social Survey, which polled tens of thousands of individuals to anonymously self-report their happiness level. Although there are many indicators present in the dataset, the ones I am focusing on are happiness, health, marriage status, working hours (including overtime), age, and income level. These indicators are the most general and can apply to all individuals.

Utilizing statistical methods to creating a logistical model for the data, we can find the probability of an individual being happy based on this large dataset.

$$\begin{aligned} \text{happiness} = & 0.399 + 1.540 * \text{healthy} + 0.504 * \text{married} \\ & - 0.0107 * \text{acturalhrs} - 0.0786 * \text{age} + 0.000762 * \text{age2} \\ & + 1.063 * \text{sixtythousand} + 1.256 * \text{ninetythousand} + 1.387 * \text{onehundredtwentythousand} \\ & + 1.357 * \text{higherincome} \end{aligned}$$

A few important notes about the model are first, that it considers slightly manipulated data to change happiness into a binary indicator (happy or not), that allows the probability of a person being happy to be measured. Exact numbers were not recorded for categorical data such as income, so the model is based on what income bracket they fall in, measured in Euros, with a value of 1 indicating that the individual belongs in the category and 0 indicating that they do not.

Lastly, “age2” represents  $\text{age}^2$ , added to uncover any nonlinear trends that correspond with happiness.

The model indicates that health contributes the most to a person’s probability of being happy, where good health constitutes a 26% higher chance on average compared to an unhealthy individual. The next highest contributor was income, but interestingly the highest bracket had a lower contribution than the second highest, showing that there is not a direct 1 to 1 relationship between the two. Age also had an interesting behavior in that the probability of being happy was higher at extreme older and younger ages compared to the middle ages. The remaining indicators contributed less but were still significant.

The visualization has two main components, an overview and a happiness calculator and comparator. The overview will utilize the model to display the indicators explored and a summary of how substantial they are by showing how a change in the indicator can increase the chance of being happy. Along with beauty and simplicity, my goal with this display is to spark curiosity so they may want to move on to the next section thinking thoughts like, “I wonder why happiness doesn’t keep increasing proportionally with higher income?” The second section allows people to create mock individuals and explore their probability based on happiness as well as compare how individuals’ probability of happiness changes for different indicators. The design allows the audience to not only see the relationship between the variables but also work to understand it themselves by modifying the variables themselves.

The complexity of the overview comes with the simplicity in formatting the data so that the point comes across easily, as well as explaining the data so that individuals can understand it easily without having to understand the more advanced statistic regression involved in the background. The main complexity that involves the second section is the ability to add

individuals to compare over different indicators such as age or working hours. This involves a dynamic dataset that the viewer can manipulate.