

# Capstone Project Proposal: Defining Happiness

*Emily McCormick*

*12/28/2018*

## **What is the problem you want to solve?**

The problem I'm trying to solve for is people's definition of happiness is defined differently across variables such as age, gender, nationality, etc. When holding all other variables constant, do we see a change in what defines happiness between males and females? Between different nationalities? What about a shift from unmarried life to married life? This is an important question to answer as people spend their entire lives trying to find happiness.

## **Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?**

My client is an app company that customers use for ideas of how they should spend their time. Additional clients could be a specific company that wants to know if their idea will truly be popular in the market (i.e. if they have a new cooking tool, is there a specific demographic they should be targeting). My client cares about this problem because it proves the validity of the product they are trying to sell. Without this, they could be giving bad ideas to the customers of what to do in their free time, or could be targeting the wrong demographic group with their new product.

## **What data are you going to use for this? How will you acquire this data?**

I found a data repository (courtesy of Data is Plural) that can be found on GitHub: <https://github.com/rit-public/HappyDB/tree/master/happydb/data>. The data comes from a team using Amazon's Mechanical Turk crowdsourcing service that asked people to describe three happy moments in the previous 24 hours or the previous three months. The dataset consists of 100,000 crowd-sourced happy moments. In addition to the free-form answers to the question, demographic information (age, gender, nationality, married/unmarried, parent/non-parent) of each respondent was also recorded. More information regarding the data collection process can be found here: <https://news.ycombinator.com/item?id=16381964>

## **In brief, outline your approach to solving this problem (knowing that this might change later).**

It will be critical to couple the raw data with a corpus of happy words. This will help me to determine where each respondent is deriving their happiness- is it sourced from something their children did? Their own accomplishments? Learning something new, or participating in a familiar tradition?

From the starting point of my corpus of happy words, I will also have to analyze if this definition of happiness changes when holding different variables constant. For example, is there a marked difference between what a 28 year-old married male from the USA sources happiness from if he has children or does not have children?

**What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.**

I will deliver both the code used for my analysis (in R), as well as a slide deck with take-aways and recommendations for companies on how to target different demographic groups.