

Galvanize Capstone 1 Proposal

Motivation

For most of my life I have been into exercising, but for most of my life I was not consistent and I never got into lifting weights. It was not until the fall of 2017 were it became important to me as a lifestyle and now I do it almost every day. My motivation was I had realized how much better I felt after a good workout. It was then that I put the pieces together; those of exercising, eating right, the various stimulus one might encounter in a day and how one reacts to said stimulus can affect your day to day wellness.

I strongly believe that one can prevent even conquer ailments such as stress, depression, and even more serious ones, with obvious exceptions, with the above remedies.

This is what guided me to finding the 'Lifestyle_and_Wellbeing_Data' on kaggle.com.

Exploratory Data Analysis

The 'Lifestyle_and_Wellbeing_Data' was gathered using 12,757 survey responses, from July 2015 until February 2020, covering the following 23 features: timestamp, fruits and veggies, daily stress, new places visited, core circle, supporting others, achievements, donations, BMI range, to do lists completed, flow, daily steps, life vision, sleep hours, lost vacation days, daily shouting, sufficient income, personal awards, time for passion, daily meditation, age, and male/female.

As far as I can tell there are no NAN values. The timestamp, age, and male/female columns are listed as objects because of their unique values. The daily stress column has one weird entry making it an object, but I am confident that the sample size is big enough that removing it will not effect my results. The final 19 columns are full of numbers.

I took the survey myself to discover my BMI and my work-life-balance score.

If you would like to take the survey yourself feel free to use this link: <http://www.authentic-happiness.com/your-life-satisfaction-score>

Interesting Hypothesis/MVP

The questions I would love to answer are four fold:

- 1) How does your age affect your work-life balance score?
- 2) How does being a male or female affect your work-life balance score ?
- 3) How do these scores compare to others who are male/female?
- 4) How do these scores compare to others of the same age bracket? The brackets used by the test are: less than 20, 21-35, 35-50, and greater than 50.

I plan to do this using Python, Pandas, Matplotlib, and Numpy. These will be my MVP.

As a MVP+ I would like to do Hypothesis Testing and look at mean values to verify Central Limit Theorem.

I would then like the CLT to lead me into analyzing different distributions of the data set as an MVP++.