https://github.com/nameyeh/project-one

Team Post-Modern Bash

Nameyeh Alam, Emre Celik, Madeleine Cieslak, Jessica Gutierrez, Stefan Sampaleanu

Background:

As a means to investigate the implications of fitness in people's' social media presence, we used Twitter's API to search the twitter activity of "fitness enthusiasts" and average users. Specifically, we analyzed tweets in an effort to see if differences exist in terms of relative "happiness". We used a VADER Sentiment analysis of tweets and twitter presence/frequency of tweets as a proxy for user's relative happiness.

Results:

Are "fitness enthusiasts" happier than the average person on twitter?

We found conclusive results confirming our hypotheses that fitness users post happier tweets. We compared the normalized mean compound value for fitness users and normal users, and the average was .183 and .106, respectively. To confirm our findings, we ran the test again but excluded neutral tweets from our analyses (where compound = 0). The difference was magnified, showing a .316 mean for fitness users and .187 for normal.

- Sentiments including neutrals

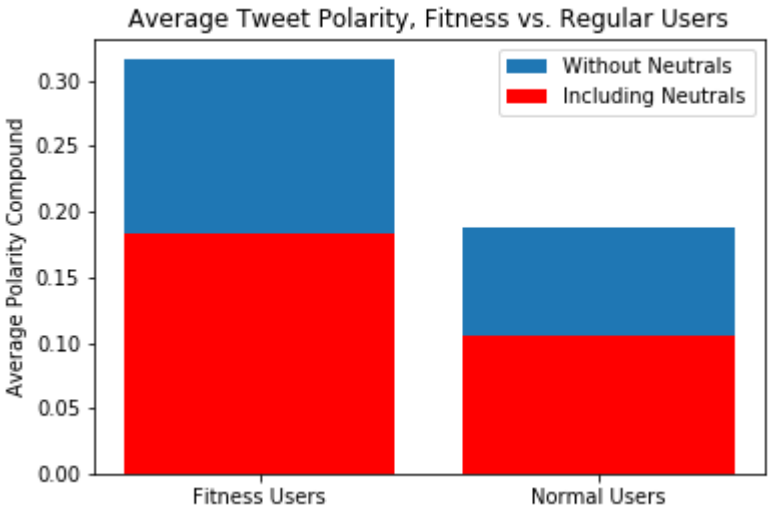
Value	Fitness User	Normal User
Compound	0.183	0.106
Positive	0.117	0.105
Negative	0.033	0.054
Neutral	0.851	0.841

- Sentiments without neutrals

Value	Fitness User	Normal User
Compound	0.316	0.187
Positive	0.207	0.185
Negative	0.058	0.094
Neutral	0.736	0.721

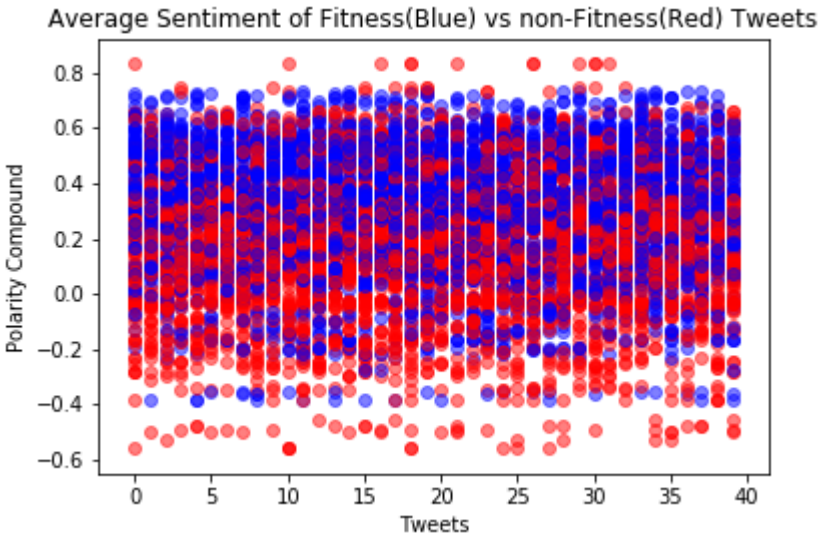
As we see, most of the difference between populations is shown in the sentiment compound, and not the positive, negative, and neutral components that make it up. This may be due to a larger range (-1;1) for compound values.

In the figure below, the red bars show mean sentiment with neutrals, and the blue shows mean sentiment without neutrals. In both cases, there is a clear difference and a t-test rejects the null hypothesis that the two groups tweet the same sentiments. This is indicated by P-Value < .05.



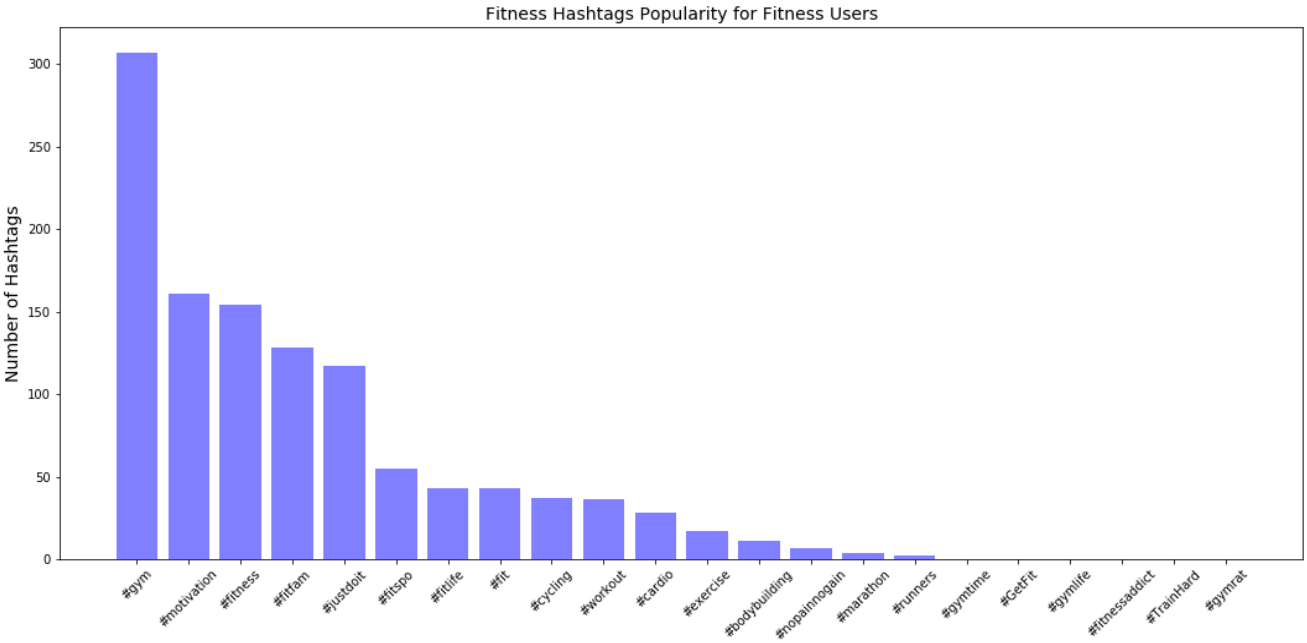
Neutrals?	N	Y
Mean Fit Sent	0.316	0.183
Mean Normal Sent	0.187	0.106
P-Value	1.023e-17	3.307e-16
T-Score	-8.740	-8.303

To create the plot below, we plotted the average sentiment of samples from both fit and normal populations (size=40) and repeated 100 times. The purpose of this plot is to demonstrate a difference in central tendency: the blue dots represent fitness users, and we can see that they occupy a higher average sentiment. The high-sentiment region above compound = .4 is dominated by blue users. Conversely, the lower half, below compound = 0 shows a stronger tendency towards normal users.

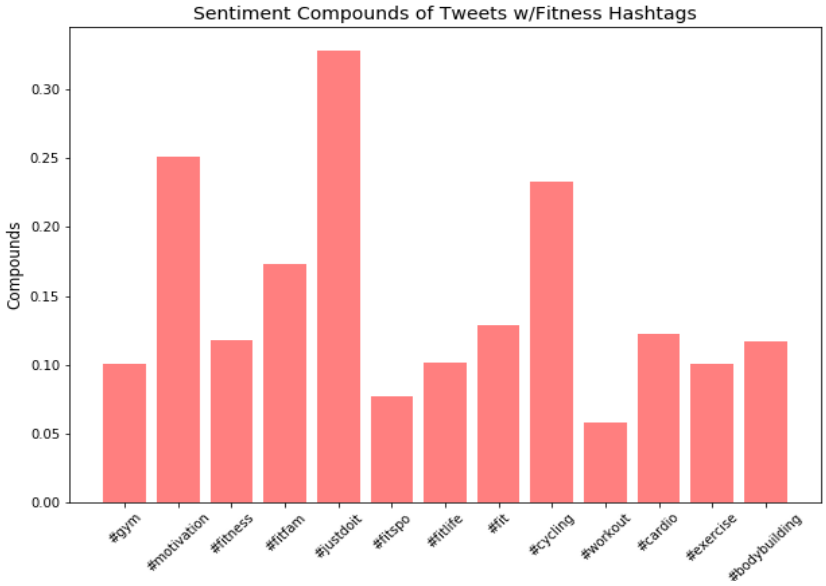


Which hashtags are popular in a random sample of “fitness enthusiasts?”

We created a subset of a hundred random Twitter Fitness users, selected one hundred tweets from each user, and filtered tweets by Fitness Hashtags. The figure below shows the popularity of the hashtags as used by our Fitness Users.

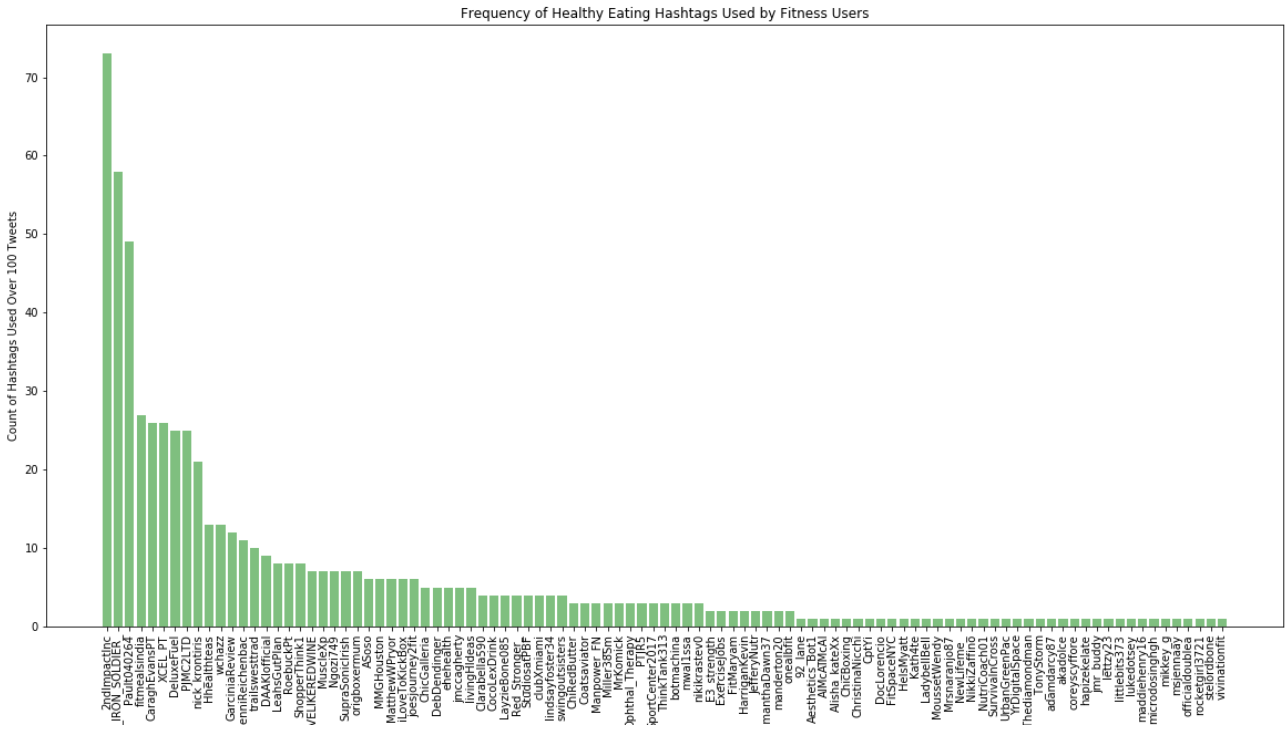


The figure below shows the mean compound sentiment score of our top 13 fitness hashtag. The hashtags #justdoit, #motivation, and #cycling have the highest positive mean scores.



Is there a link between “fitness enthusiasts” and healthy eating?

We defined healthy eating hashtags as EatClean, FitFood, GlutenFree, HealthyEating, HealthyRecipes, Nutrition, Protein, CleanEating, FitnessFood, Diet, PlantBased. We looked at past 100 tweets for each of our users, and counted how many times a user tweeted using at least one of the hashtags.



We can’t conclusively say that fitness enthusiasts are also into healthy eating, but there is a higher probability that they will tweet about nutrition over a random group of “normal” Twitter users. In our analysis, not one of the users in the normal group tweeted any hashtag related to healthy eating.

In comparison, 100 out of our 500 fitness users used healthy eating hashtags. Out of those 100 users, they tweeted about healthy eating on average 6% of the time. Although there isn’t a strong correlation between fitness and healthy eating, we found it interesting that there was such a clear difference between our two user groups.