New Twitter Trends

Firas Ben Othman, Omar Boujdaria

Abstract— In this report, we will mention the trends that we chose, define these trends, describe how we extracted the tweets, and talk about the datasets used. We will also explain how we performed the sentiment analysis for our tweets as well as the preprocessing done to perform sentiment analysis and to extract tweets. Finally, we will explore the results and draw conclusions.

INTRODUCTION

In the last few years, we have seen and heard of many new lifestyle trends/practices such as veganism, minimalism, meditation and so on... that completely changed the life of the people advocating them.
Some examples of this are Hollywood actor Terry Crews who said that ever since he started practicing intermittent fasting he looks and feels better than he did when he was 22 years old, or Janette Murray who miraculously cured her cancer by going on a raw vegan diet and ran a marathon everyday in 2013 at 64 years old.

However, everyone's body and mind are different. So the question is : Does the general public respond well to these practices too? Or do they mostly suffer from negative effects?

In our project, we decided to put that to the test. We extracted tweets talking about trends/practices of our choice and performed sentiment analysis on them to get an idea about how they are perceived and compared them to each other to understand which practice has impacted people lives for the better the most, as well as, some additional analysis that will be shown in the rest of the report.

I. DATASETS

A. Swiss-tweet

Brief description :
Swiss tweet is a dataset containing tweets posted in Switzerland in 390 files in json format.
Information in dataset :
The useful information to our project contained in these files are :

- Tweet text
- Language of the tweet
- Date of post
- Sentiment
- Gender

Note that some sentiments are missing from the dataset.
General Statistics :
There are 10,828,070 tweets in our dataset, all located in Switzerland.

B. Tweets-leon

Brief description :
This dataset is a big text file containing tweets and some additional information about each tweet.
Information in dataset :
The useful information to our project contained in these files are :

- Tweet text
- Language of the tweet
- Date of post

General Statistics :
The dataset contains 33'723 files containing billions of tweets posted in the span of 2011 to 2016.

II. TRENDS ANALYZED:

For our project, we chose to analyze the following practices :

- **Intermittent Fasting**
  Practicing intermittent fasting consists of eating all the daily calories within a small window of time (e.g 8 hours) during the day, usually, in the evening. One can drink water or black coffee during the fasting periods.

- **Veganism**
  Veganism is the practice of eating and using no animal products at all.

- **Paleo Diet**
  The paleo diet / ketogenic diet consists of eating only food presumed to be available to the "cavemen" which consists strictly of meat, fish, nuts, fruits and vegetables.
  This diet excludes dairy products, oils, grains and cereal.

- **Meditation**
  It isn't easy to define meditation by one's own words, so here is the definition on a Wikipedia page.

Meditation is a practice where an individual operates or trains the mind or induces a mode of consciousness to allow the mind to engage in peaceful thoughts. Meditation is often used to clear the mind, reduce stress, promote relaxation or train the mind.It may be done while sitting, repeating a mantra, and closing the eyes in a quiet environment.

- https://en.wikipedia.org/wiki/Meditation

- **Minimalism**

Minimalism is the practice of having as few possessions as possible. (e.g. One room appartment, very few clothes, no car ...)

III. EXTRACTING TWEETS

Extracting target tweets was done by creating a list of keywords for each trend in the three main languages in Switzerland and English.

A. Preprocessing

To increase matching probability using keywords, the tweets had to be preprocessed.
The preprocessing was done as follows :

- Turn all letters of the tweet to lowercase.
- Delete all spaces.
- The special letters of French, German and other languages such as "é" or "ä" were transformed to their base letter (meaning e and a respectively for the letters mentioned as examples)

B. Results For Swiss-tweet :

After the tweets were extracted, we took a look at how good and accurate the results are.
- **Veganism** :
The results for veganism are good. The vast majority of the tweets extracted are not out of context.
The total number of tweets extracted is 5908.
- **Intermittent Fasting** :
The results of extracting tweets about intermittent fasting from swiss-tweet are bad as it only resulted in returning 3 tweets.
- **Paleo Diet** :
There are 91 tweets for Paleo diet tweets extraction. This number is not big at all. In addition, quite a few of these resulting tweets are out of context. Most of these tweets are in relation with the Paleo festival.
- **Minimalism** :
The number of resulting tweets about minimalism is 423. But almost all of them are out of context

- **Meditation** :
There were 6490 results for tweets about meditation. However, by looking at the resulting tweets, we found that some of them are tweets of a radio station called meditation and some other tweets are ads for guided meditation sessions.
But, the number of tweets that are out of context is not catastrophic.

C. Results For Tweets-leon :

By using the same method on the tweets-leon dataset, we extracted tweets about the five subjects and took a look at the results.
- **Veganism** :
We got 174,282 tweets for veganism. Most of them are accurate, however, because of the structure of the tweets-leon dataset, some results are from the username even if the tweet posted doesn't talk at all about veganism.
- **Intermittent Fasting** :
We got 723 tweets for intermittent fasting. Most of them are accurate and the number of results is decent as well. This was encouraging considering how bad the results were for swiss-tweet.
- **Paleo Diet** :
The number of results for paleo diet is 225,402.
It is not possible to verify too many of the tweets. But, by looking at a good number of them, the results seem pretty accurate.
- **Minimalism** :
The result for minimalism is huge but very inaccurate. In fact, we couldn't find one tweet talking about minimalism as we intended it to mean. Maybe it is not very talked about after all.
- **Meditation** :
We found 5651 about meditation, and the results were quite good too.

D. Decreasing Inaccuracy

To decrease the inaccuracy for some tweets extracted, we decided to choose a list of negative keywords for each tweets set that has many out-of-context results.
These negative keywords indicate a common mistake in tweets extraction. For instance, a negative keyword for the paleo diet could be 'paleofestival' since there many tweets that talk about the paleo festival instead of the paleo diet.

As for minimalism, we decided it was better to just drop it instead of analyzing this trend using a bad set of tweets.

IV. RUNNING SENTIMENT ANALYSIS

Even though Swiss-tweet dataset already has sentiment for most tweets, the same Sentiment Analysis algorithm was run for both tweets from swiss-tweet and tweets from tweets-leon, for the sake of fairness.

A. Preprocessing for sentiment analysis

As the pre-trained sentiment analysis algorithms work best for sentences in English, we decided to translate all non-English tweets to English as best as possible.

To do so, we used the API googletrans, which is a free Google Translate API for Python.

Also, for a reason still unknown to us, tweets that contain web page links produce a $JSONDecodeError$. That is why, we deleted all links from tweets before proceeding to translating tweets to English.

B. Sentiment Analysis

The library used for Sentiment Analysis is *SentimentIntensityAnalyzer* from the package *nltk.sentiment.vader* . This library is perfect for our project, since it is known to work well for social media texts in English.

The function *polarity_scores* from *SentimentIntensityAnalyzer* returns probabilities of a sentence being positive, negative or neutral.

In order to reduce the number of neutral results efficiently, we selected a threshold of $Prob_{NEUTRAL} \geq 0.70$ to actually give to a tweet a sentiment of neutral. Otherwise, the sentiment of a tweet takes the maximum of POSITIVE or NEGATIVE in terms of probability.

V. SENTIMENTS

A more in-detail comparison between trends as they are viewed in Switzerland and the rest of the world will be done in section Data exploration.

In this section, we will be looking at sentiments from one dataset at a time.

A. Swiss tweets

We can see, from Fig.1, that the tweets about meditation are overwhelmingly positive and contain very few negative tweets.

On the other hand, sentiments about the paleo diet are way more negative than positive, we could assume that people didn't do well on a paleo diet compared to the other two trends, but the number of tweets about paleo diet is very small compared to the number of tweets about meditation and veganism. So, we can't make that claim.

The sentiments of tweets about veganism are more so positive than negative, but we can say that meditation takes the win within Switzerland.

B. Tweet leon

From Fig.3, we can see that most of the trends get positive sentiments (if we don't consider neutral tweets) and don't have a lot of negativity. But we can see that veganism has a slight edge over the other trends in terms of positivity and meditation has the least negative sentiments per tweet.

The paleo diet gets mostly bad sentiments in the tweets posted by users worldwide similarly to those posted in Switzerland.
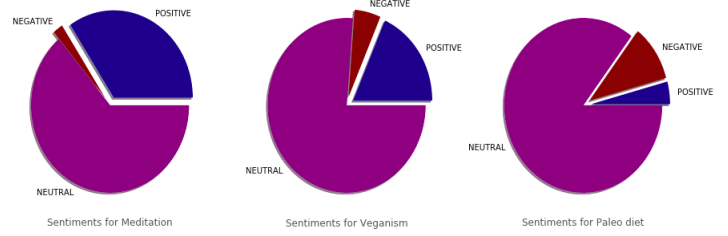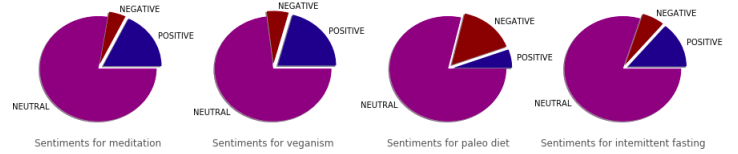

Fig. 1. Sentiments worldwide

Fig. 2. Sentiments by year for each trend in Switzerland

VI. EXPLORING RESULTS

For most of the next plots, we won't represent neutral sentiment, since it doesn't represent any kind of user bias for most cases.

That is of course with the exception for section C.2 where we used neutral sentiments to see the amount of interactions of users with the tweets.

A. Comparing Swiss tweet and Tweets leon Sentiments

As We can see in Fig. 3, the trends veganism, intermittent-fasting and meditation all get positive sentiments.

The sentiments in Switzerland for meditation is significantly more positive and less negative than that of the tweets in the world.

The Paleo diet on the other hand doesn't do well in Switzerland and in the world which suggests that the people that experimented with this trend had many complaints about it.

Intermittent fasting, unfortunately, has very few results by extracting it from Swiss Tweet so we preferred not to represent it. But it gets generally good sentiments, comparable to veganism and meditation.

Finally, veganism also got good sentiments and it even had a slight edge on meditation in terms of positive sentiments.

B. Grouping by year

The grouping by year, unfortunately, could only be done for tweets extracted from tweets leon dataset. This was because, we had only result about tweets posted in 2016 for Swiss tweet dataset.

As we can see in Fig.4 , the graphs show that there aren't significant jumps or falls in popularity of the practices, except in the case of the paleo diet.

The paleo diet seems to have been talked about a lot in 2011, and hated on a lot. In fact in 2011 we found a percentage of negative tweets about Paleo diet of over 60% (The bar was cut short to show the other bar plots alongside it clearly with same y scale, look at the notebook of the project to
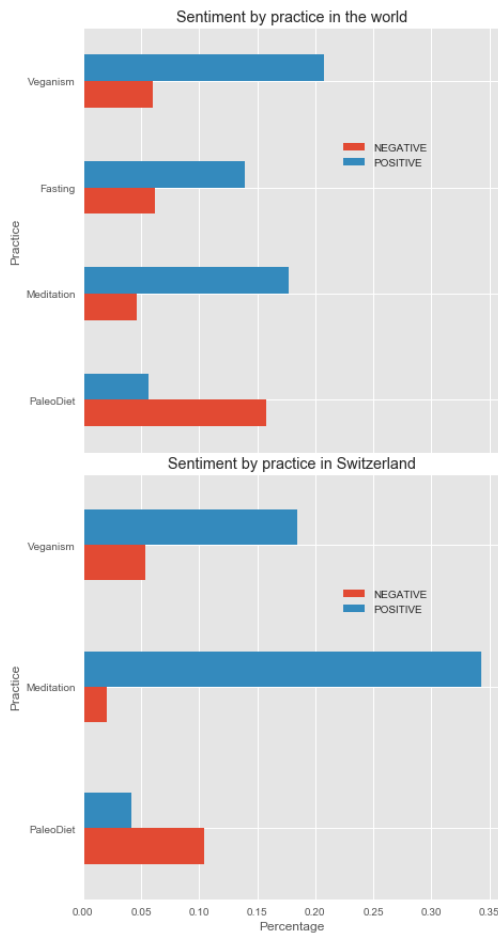
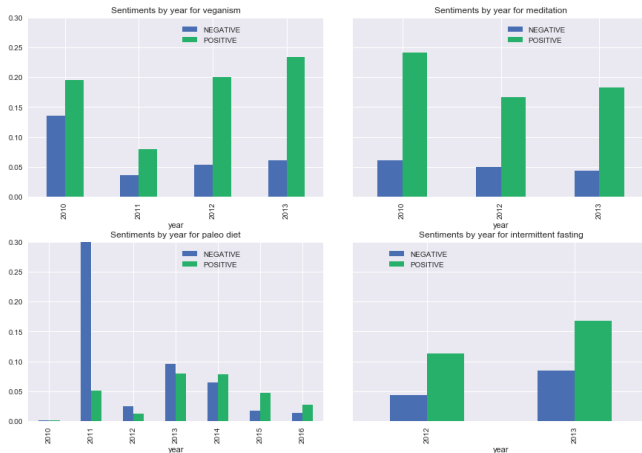Fig. 3. Sentiments by trend in Switzerland and worldwide

Fig. 4. Sentiments by year for each trend in Switzerland

see full plot). As for the rest of the years, the difference between sentiments is very small and their fraction values is very small as well. This means most of the sentiments were neutral.

Veganism, meditation and intermittent fasting have had very small differences through the years, and their sentiments are mostly positive.
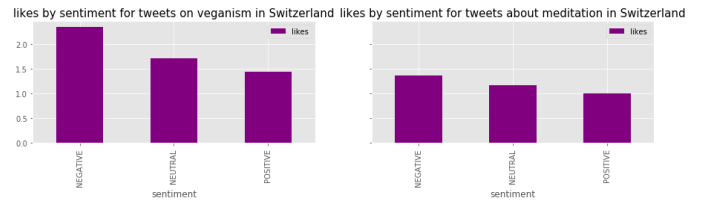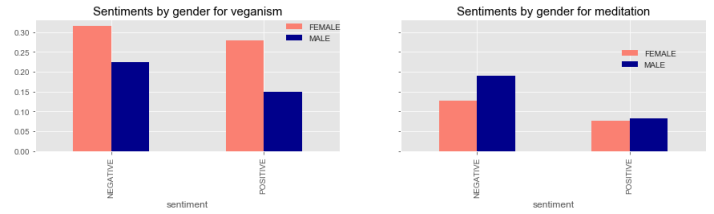

Fig. 5. Sentiments grouped by likes

Fig. 6. Sentiments grouped by gender

C. Further exploration

In addition to grouping results from the tweets leon dataset by year, we grouped the results from Swiss tweets by user gender and by likes.

However, the tweets about paleo diet extracted from swiss tweets had lots of NAN values as number of likes and $UNKNOWN$ values for gender. That is why, we couldn't include paleo diet in these explorations.

1) grouping by likes: We see that from in Fig.5, veganism gets , in general, significantly more likes. Which may suggest that this topic gets more reactions from users.
For both cases, the negative tweets get more likes than neutral and positive tweets. This shows that people love to hate on perfectly good practices (as concluded from sentiments graphs).

2) grouping by gender: As show in Fig.6, veganism is more talked about amongst female users in general as they are more numbered both in positive and negative tweets.
As for meditation the ratio is much closer, but male twitter user talk about it slightly more.

CONCLUSIONS

The Paleo diet is not very popular in Switzerland or in the rest of the world. Although, it is presumed to work well for humans because it consists of eating as we evolved to eat. This assumption doesn't seem to be true after all.
All of veganism, intermittent fasting and meditation get fairly positive sentiments. You couldn't go wrong with experimenting with any of the three. They are all associated with increased energy, euphoria according to some tweets we found.And this is backed up by the general results we found. In terms of evolution by year, it looks like the paleo diet was talked about in 2011 more so than other years. However, the difference in popularity in the other three trends between years is negligible.