Social Data Science Project – Sentiment in vegan-related tweets
by Andreas Opedal & Charlotte Out

Introduction

Veganism can be defined as "a way of living which seeks to exclude, as far as is possible and practicable, all forms of exploitation of, and cruelty to, animals for food, clothing or any other purpose.", according to the vegan society (2014). It includes following a so called "plant-based diet" thus excluding foods such as meat, fish and insects, similarly to vegetarians, but also eggs, dairy, honey and all other products in which animals were involved in the production process. Often, the reasons to adapt a vegan diet are divided into three categories; health reasons, animal rights and environmental preservation (Greenebaum, 2018).

Traditionally, vegans have often been framed in negative light in mainstream media. According to Cole and Morgan (2011), in 2007, 74.3 percent of the articles related to veganism in the United Kingdom daily and Sunday newspapers were classified as having negative sentiment, while only 5.5 percent of these articles were classified to be positive (thus 20.2 percent were neutral). In particular, in the British newspaper and tabloid "The Daily Mail", 6.5 percent of the articles concerning veganism had positive sentiment and 89.1 percent had negative sentiment. The aforementioned authors found that a large number of articles aimed to ridicule veganism, described veganism as a fad and characterized vegans as oversensitive.

However, recently, there has been a change in media portrayal and sentiment towards veganism (Lundahl, 2018). Vegan diets have experienced an increase in popularity, predominantly among teenagers and youth, and especially females (Craig, 2009). This change could be influenced by a number of factors. Firstly, there is a growing body of research relating plant-based diets to positive health outcomes (Craig, 2009). It is related to lower body mass index, modestly lower blood pressure and lower cholesterol levels (Craig, 2009). Secondly, there is a growing body of research highlighting the effect of veganism on environmental sustainability. Crucial is that much of this research is not exclusively communicated in academic papers, but also to more accessible news sources such as newspapers and online articles. Examples of this are articles such as "Avoiding meat and dairy is the 'single biggest way' to reduce your impact on Earth" and "Plant-based diet can fight climate change" by the BBC. In addition to this, popular documentaries promoting vegan diets such as "Cowspiracy", "What the health" and "Game Changers" have gained much media attention. This has led to veganism becoming more "mainstream" (The Guardian, 2018). The year 2014 for instance was declared as "the year of the vegan" as many celebrities such as Beyoncé, Jay Z and Gwyneth Paltrow associated themselves with the diet (Rami, 2014). In this period there has been a significant increase in the number of people following a vegan diet. It is estimated that this number rose from 150,000 in 2006 to 542,000 in 2016 in the UK, which is a 350% increase (Lundahl, 2018).

Over these years the sentiment towards veganism also changed drastically. It was found that in 2014, only 14.7 percent of the articles related to veganism in "The Daily Mail", were classified as negative, and 76.7 percent of the articles as positive, providing a sharp contrast with the findings of Cole and Morgan 7 years earlier. This change in sentiment is highly remarkable. However, these results are limited to the sentiment towards veganism in the UK.

In this study, we aim to identify how the attitude towards veganism differs from country to country, and if there is a relationship between country-level statistics and the attitude towards veganism. We use the sentiment of vegan related tweets on Twitter as a proxy for a country's overall attitude to veganism and investigate the correlation between this and GDP/capita, Future Orientation Index (FOI), Global Innovation index (GII), Human Development Index (HDI) and the Climate Change Performance index (CCPI). In doing so, we hope to increase the understanding of what factors might influence a population's sentiment towards veganism, as well as provide new insights. More specifically, we aim to answer the following research question:

- Is there a positive correlation between the sentiment of tweets related to veganism and economic wealth, future orientation, innovativeness, welfare and performance in combatting climate change?

Our expectation is that there is a positive relationship between the sentiment of tweets (higher value meaning more positive) and all of the investigated factors. Motivation is provided below:

- GDP/capita is a measure of economic activity in a country but is often used as a proxy for standard of living. In countries with a higher standard of living people are arguably more open to adopt to new eating habits, having the financial means to be able to do so. People would thus be more open to vegan diets.

- It is well established that adopting aspects of a vegan diet such as eliminating meat consumption reduces the carbon footprint. This in turn has a positive effect on combatting global warming. We argue that countries with a higher level of future orientation would take more action towards combatting climate change (even if this action has little significance large scale). We therefore hypothesize that countries with a high Future Orientation Index also have a more positive sentiment in tweets concerning veganism.

- As for the Global Innovation Index, there is currently a high level of innovation in the FoodTech industry to satisfy the increasing demand in more sustainable food, examples being oat-milk and plant-based beef. We suspect that companies that innovate in this area have a high demand in their domestic market, thus motivating our expectation.

- The Human Development Index takes per capita income as well as life expectancy and the duration of education into account. Per capita income was argued for above by GDP/capita. Level of education is also believed to have a positive correlation with vegan sentiment – the positive impacts of vegan diets might be more unknown to less educated populations.

- As mentioned above vegan diets have a positive impact on the environment. It therefore seems likely that countries with a higher Climate Change Performance Index should have a more positive sentiment towards veganism.

Theoretical Framework

Sentiment analysis

Sentiment analysis or opinion mining is a set of tools to identify and extract opinions and sentiment from text. Sentiment analysis can happen at three main classification levels (D'Andrea et al., 2015)

1) Document level: Classifying the document as having a positive or negative sentiment
2) Sentence level: Classifying the sentiment expressed in each sentence
3) Word level: classifying the sentiment of each word

Furthermore, sentiment analysis approaches differ and can roughly be classified in the following three ways:

1) Machine learning approaches or supervised sentiment analysis
2) Lexicon-based approaches or unsupervised sentiment analysis
3) Hybrid approaches

Lexicon based approaches do not require prior training, as it uses a predefined list of words or a vocabulary, where each word has a sentiment value assigned to it. Examples of lexicon-based tools are

Linguistic Inquire and Word Count (LIWC), SentiStrength, Syuzhet, VADER, Opinion Finder and Multi-Perspective Question Answering (MPQA), which were all introduced in the course. Advantages of these approaches is that they are generally simple to implement, and they have a large coverage and recall. However, they are not easy to customize for a particular context.

Supervised Machine learning approaches require a model to be trained on data before it can be used. It often requires learning a so called "embeddings" where words or phrases are mapped to vectors in a typically lower dimensional vector space. A characteristic of a (good) embedding is that similar words or phrases, should be close together in this vector space. When this representation is created, one can use several classification algorithms to extract the sentiments of the tweets. Examples of such classification algorithms are support vector machines, logistic regression, and neural networks (D'Andrea et al., 2015).

Correlation coefficients

The Pearson correlation coefficient is a measure of linear correlation between two variables. It takes values between +1 and -1, where +1 represents a total positive linear correlation, 0 represents no linear correlation and -1 a total negative linear correlation. It is defined as the covariance between two variables X and Y, divided by the product of the standard deviations of X and Y.

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Spearman's correlation coefficient is a measure of the statistical dependence of the ranking of two variables. It is therefore a non-parametric measure. Spearman's correlation assesses whether there is a monotonic relationship between the variables. A Spearman correlation of +1 or -1 is obtained when there is a perfect monotonic relationship between the two variables X and Y we consider.

$$r_s = \rho_{\mathrm{rg}_X, \mathrm{rg}_Y} = \frac{Cov(\mathrm{rg}_X, \mathrm{rg}_Y)}{\sigma_{\mathrm{rg}_X} \sigma_{\mathrm{rg}_Y}}$$

Here rg denotes that it is applied to the rank variables.

Permutation tests

Permutation tests are used to test hypotheses. More concretely, they quantify how plausible it is that our measurements or statistics are produced by chance under a null hypothesis. The general idea is to use permutations of the dataset to destroy the relationship that is to be tested.

A permutation tests consists of three components:

1) A test statistic, for example a correlation coefficient
2) A null hypothesis, usually what would happen under complete random assignment and an alternative hypothesis, the hypothesis that we are testing
3) A permutation set with N random reshufflings of the data, for which the null hypothesis holds

We compute the given statistic on the N random reshufflings of the data to obtain the permutation distribution. This is then compared to the observed non-permuted test statistic.
After this the p-value is computed, which is the proportion of values in the permutation distribution that are as extreme as or more extreme than the test statistic of the data. If the p-value is below a set Type I error probability, the null hypothesis is rejected in favor of the alternative.

Methodology

We started off by collecting the data with the help of a developer account from Twitter. We wrote a search query including the term "vegan" and additionally also the terms "plantbased" and "dairyfree" since they were deemed to be closely associated[1]. Retweets were also included as they are argued to be equally indicative of the overall sentiment. The search query was executed in batches of 10,000 tweets (extracting the 10,000 most recent tweets within a week of passing) over a period of several weeks from the middle of January to the middle of March, 2020. This resulted in a data set of around 200,000 vegan-related tweets.

The vast majority of tweets did not include a so called "geo-tag", which would have made the tweet's country of origin directly visible in the data. To extract the country we instead geocoded the location field associated with the Twitter profile that produced the tweet. This part of the analysis was done in Python, using the geocode service GeoNames. GeoNames takes the location as an input and then returns a complete address, including the country. The countries we chose to focus on were limited to the OECD countries plus Brazil, India, Russia and South Africa. We then wrote a function that using a dictionary-based approach extracts the aforementioned countries if they occur in the address returned by GeoNames, and inserts this as a new column in the dataset. The large volume of data that was initially collected got reduced to around 25,000 observations after this step. This large difference was mainly due to three reasons:

1. Not all Twitter accounts had a location set and only a subset of those included an actual location.
2. The number of requests that could be made from GeoNames within a given timespan was limited.
3. The accuracy of GeoNames was satisfactory, but not 100%.

After data extraction and pre-processing the sentiment of the vegan tweets could be analyzed. The open source sentiment analysis tool VADER was selected for this purpose. Due to limitations in language support of the sentiment analysis tools available for R we had to restrict the analysis to tweets made in English. This restriction did not however significantly reduce the number of countries that could be included in the analysis. The average vegan tweet sentiment was then computed for each country with a number of English tweets above 100. We then computed the Pearson and Spearman empirical correlation coefficients between this variable and GDP/capita, HDI, GII, CCPI and FOI which were collected from the World Bank, the United Nations Development Program, the Global Innovation Index website, the Climate Change Performance Index website and Google Trends, respectively. Permutation tests with 1000 simulated values were then performed for each of these empirical correlation coefficients, with the null hypothesis of no correlation and a Type I error probability of 0.05.

Analysis

The correlation coefficient between the average vegan sentiment of a country and its GDP/capita was found to be 0.2817 using the Pearson correlation coefficient. As the Pearson correlation coefficient is known to be sensitive towards outliers, we also computed the more robust Spearman's rank correlation coefficient, to add validity to our analysis. This was found to be 0.1929 for vegan sentiment and GDP/capita. The distribution over the Pearson correlation coefficients obtained in the permutation test is shown in Figure 1. The p-value for the Pearson correlation and Spearman correlation were found to be 0.1558 and 0.2537, which under Type I error being specified at 0.05 provides no evidence against the null hypotheses. We therefore do not reject the null hypothesis of no correlation.

[1] As we initially also planned to analyze tweets made in other languages than English, we translated the word "vegan" to a number of languages and included those terms in the search query as well.

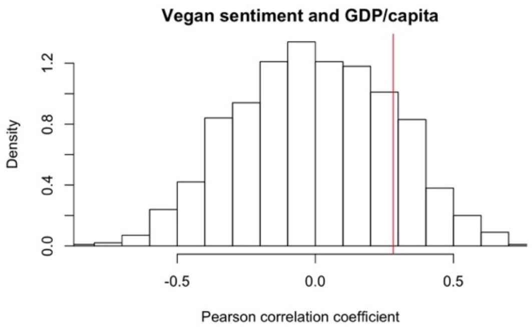

Figure 1 Histogram over simulated values of the Pearson correlation coefficient between vegan sentiment on Twitter and GDP/capita. The red line indicates the observed value.

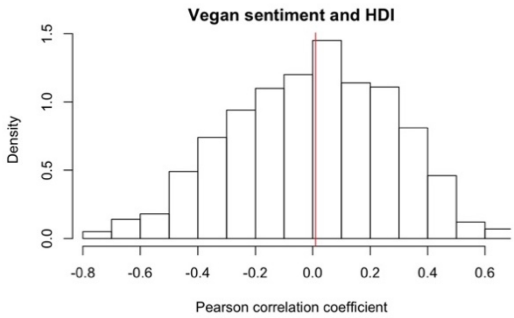

Figure 2 Histogram over simulated values of the Pearson correlation coefficient between vegan sentiment on Twitter and HDI. The red line indicates the observed value.

The Pearson and Spearman correlation coefficients between average vegan sentiment and HDI per country were found to be 0.0103 and 0.2395 respectively. The distribution of the Pearson correlation coefficients obtained in the permutation tests is shown in Figure 2. The p-values were found to be 0.5235 and 0.1788, which under Type I error specified at 0.05 is again not significant, and we therefore do not reject the null hypothesis.

Between vegan sentiment and GII the Pearson and Spearman coefficients were found to be 0.1605 and 0.2500. The distribution over the correlation coefficients is shown in Figure 3. The correlation coefficients between the average vegan sentiment and CCPI and average vegan sentiment and FOI were found to be -0.0582 and 0.0494 respectively using the Pearson correlation coefficient and -0.2357 and 0.1468 using the Spearman correlation coefficient respectively. Their distribution over the correlation coefficients are shown in Figure 4 and Figure 5 respectively. For the last three correlations we obtained p-values which were also not significant under Type I error specified at 0.05 (0.2997, 0.6044, 0.4326 respectively for Pearson correlation coefficient and 0.1818 0.8082, 0.3197 respectively for Spearman correlation coefficient).

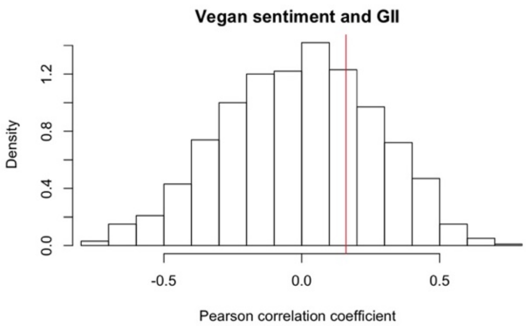

Figure 3 Histogram over simulated values of the Pearson correlation coefficient between vegan sentiment on Twitter and GII. The red line indicates the observed value.

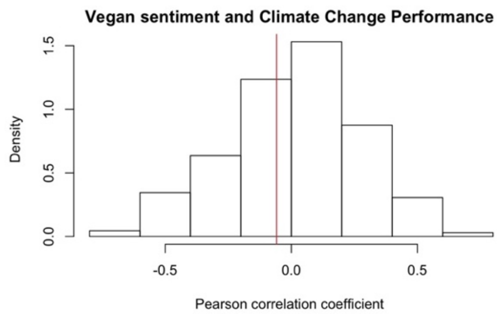

Figure 4 Histogram over simulated values of the Pearson correlation coefficient between vegan sentiment on Twitter and CCPI. The red line indicates the observed value.

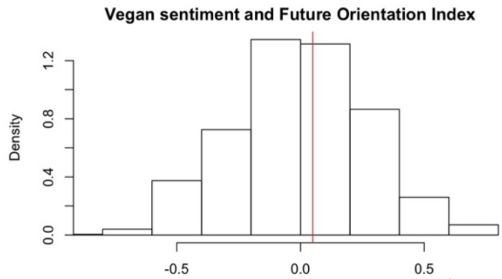

Figure 5 Histogram over simulated values of the Pearson correlation coefficient between vegan sentiment on Twitter and FOI. The red line indicates the observed value.

Discussion

Against our expectations, we found all of our statistics (GDP/capita, HDI, GII, CCPI and FOI) to not be significantly correlated to the average sentiment towards veganism on Twitter. This could mean that the attitude towards veganism is not correlated to these factors, but it could also be due to a number of limitations in this study. Aggregating the sentiment towards veganism in tweets by taking the average might not give a good representation of the sentiment towards veganism in a country. It could for example have been the case that the sentiment of most of the tweets are positive, but that there are a few very negative tweets bringing this down. To test this we also checked the median of the sentiment towards veganism of the countries we focus on, and computed the correlations with this. This however also resulted in low correlation coefficients (all below 0.25). It could then be the fact that an overall positive attitude towards vegan diets might not mean that tweets associated with veganism are all positive. There might for instance be tweets complaining about why not more people follow the vegan diet, reactions to animal abuse or the like, which would carry negative sentiments. It could also be that people with negative opinions on topics carry "larger voices", in the sense that they are more likely to share their opinion on social media. Taking these possibilities into account, sentiment in vegan related tweets might not be the best proxy for a country's overall attitude towards a vegan diet.

Another limitation is that the analysis had to be restricted to tweets made in English. As mentioned above, this did not significantly reduce the number of data points – several countries with a non-English native language had a large number of tweets made in English. However, it might be that the tweets made in English do not properly represent the overall sentiment. For example, it could be that only more educated individuals write English tweets, which would create a selection bias in our data set.

Furthermore, all of these tweets were collected during the beginning of the COVID-19 crisis. This had a large impact on the topics discussed on Twitter and on the overall Twitter environment, which probably resulted in a sample that is not fully representable of the ordinary. There might for example be less vegan related tweets than otherwise or that more tweets overall are more negative in sentiment.

We can however only hypothesize about why the results in this study were not as expected. All we can conclude is that there is no association between the sentiments of vegan-related tweets on Twitter with neither economic wealth, future orientation, innovativeness, welfare nor performance when it comes to combatting climate change.

References

Cole, M., & Morgan, K. (2011). Vegaphobia: Derogatory discourses of veganism and the reproduction of speciesism in UK national newspapers. *British Journal of Sociology*, *62*(1), 134–153. https://doi.org/10.1111/j.1468-4446.2010.01348.x

CRAIG, W. J. (2009). file:///Users/marieketheil/Desktop/epic_veg.pdfHealth effects of vegan diets. *American Journal of Clinical Nutrition*, *89*, 1627–1633. https://doi.org/10.3945/ajcn.2009.26736N.Am

D'Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, Tools and Applications for Sentiment Analysis Implementation. International Journal of Computer Applications, 125(3), 26–33. https://doi.org/10.5120/ijca2015905866

Greenebaum, J. (2012). Veganism, identity and the quest for Authenticity. *Food, Culture and Society*, *15*(1), 129–144. https://doi.org/10.2752/175174412X13190510222101

Lundahl, O. (2018). Dynamics of positive deviance in destigmatisation: celebrities and the media in the rise of veganism. *Consumption Markets and Culture*, *0*(0), 1–31. https://doi.org/10.1080/10253866.2018.1512492

The Guardian (2018). 'The unstoppable rise of veganism: how a fringe movement went mainstream'. The Guardian, 1 April [Online]. Available at https://www.theguardian.com/lifeandstyle/2018/apr/01/vegans-are-coming-millennials-health-climate-change-animal-welfare