Climate Change Twitter Buzz and Recycling

Team Datasaurs (kfriis, litofish, ktan7, wjurayj)

Hypothesis

Social media is often used to raise awareness about issues such as climate change, often with the intent of spurring action from followers. We sought to investigate whether posts on social media actually lead to more climate conscious behavior. Specifically we tested the following hypothesis: greater Twitter engagement on climate topics leads to more recycling.

Data

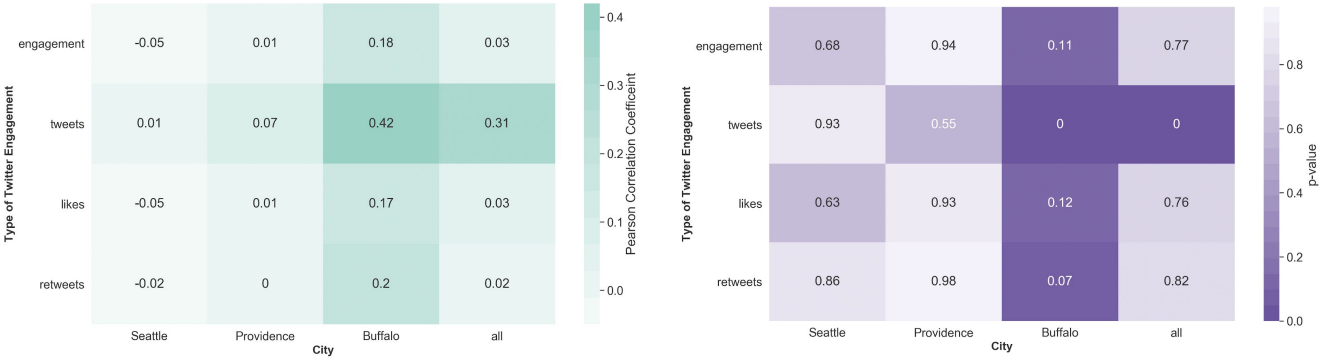
We collected Twitter data and recycling data from Seattle, Providence, and Buffalo. We used the Twitter API to query tweets mentioning 'climate,' posted between 2013-2019, and geotagged in one of the three cities. The raw recycling data was obtained from open datasets published by each city's government, which we used to calculate monthly recycling diversion rates. We cleaned and combined our data into a table with the following columns: year, month, num_tweets, likes, retweets, engagement (likes + retweets), and diversion_rate.

From a sample of 322 tweets, we found that 19% of our queried tweets were false positives that either used "climate" unrelated to climate change or were denying climate change. We also found the recycling diversion rates followed a cyclic pattern and spiked every year during the holiday season (Nov, Dec, Jan), which is due to an increased amount of waste and recyclable material produced during the holidays.

Findings

Claim #1: Higher recycling diversion rates are associated with a higher number of tweets about climate change in (1) Buffalo and in (2) Buffalo, Seattle, and Providence aggregated.
Support for Claim #1: We looked at the Pearson correlation coefficients and their *p*-values for recycling diversion rate and each metric we had collected from Twitter. (1) We found that the correlation coefficient of number of tweets versus diversion rate in Buffalo was 0.42 with a *p*-value of 0.00. Using a significance level of 0.05, we find this to be a significant positive correlation. (2) We also found that the correlation coefficient of number of tweets versus diversion rate for all the cities aggregated was 0.31 with a *p*-value of 0.00. We find this to also be a significant positive correlation. For all the other relationships and cities there were no significant correlations.

Pearson Correlation Coefficients (left) and p-values (right) of Twitter Features



Claim #2: Holiday season has a greater influence on recycling diversion rates than the number of tweets about climate.
Support for Claim #2: From our initial data exploration, we knew that the recycling rates spiked during holiday seasons, so we introduced a dummy variable for whether or not it was the holiday season (Nov, Dec, Jan). After controlling for this, there was little correlation between the number of tweets and recycling diversion rate (coefficient=0.0132) whereas holidays showed a much stronger correlation (coefficient=3.95).

Aggregate	Coefficient	*p*-value
Holidays	3.9479	0.000
Number of Tweets	0.0132	0.001
Engagement	-8.701e-05	0.294
Intercept	18.3204	0.000

Claim #3: In Buffalo, the city's recycling campaign has had a greater influence on recycling diversion rate than the number of tweets about climate.
Support for Claim #3: We found that since 2015, the city of Buffalo has promoted a campaign to bring recycling rates above the national average of 34%. To account for this, we created a dummy variable for whether or not there was a recycling campaign going on. After adding the campaign and holiday season variables to our multiple linear regression, we found no significant correlation between diversion rate and Buffalo's Twitter metrics (p-value=0.392). This is consistent with the findings for the other cities.

Buffalo	Coefficient	*p*-value
Holidays	6.5074	0.000
Campaign	7.4568	0.000
Number of Tweets	0.0902	0.329
Engagement	0.0007	0.808
Intercept	8.8866	0.000

Claim #4: Greater twitter engagement does not cause higher recycling diversion rates.
Support for Claim #4: After introducing additional variables in our regression (claims #2 & #3), the initial relationships we saw when looking at the Pearson correlation coefficients (claim #1) proved to be insignificant. This leads us to conclude that our hypothesis does not hold, and that Twitter engagement on climate topics does not lead to more recycling. Rather, the holiday season and city recycling campaigns seem to be bigger drivers of recycling rates than Twitter engagement on climate topics.