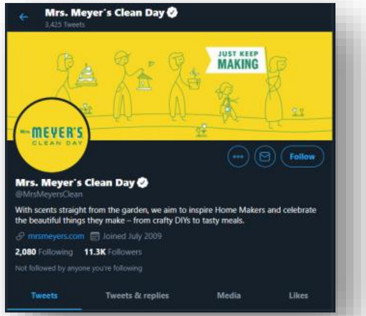
SENTIMENT ANALYSIS PRIMER:

HOUSEHOLD CONSUMER PACKAGED GOODS (CPG)

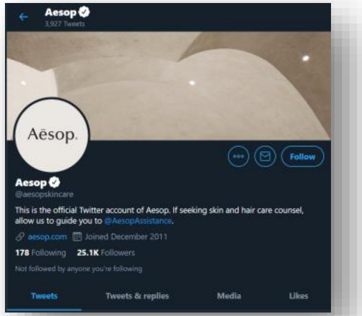
ANALYSIS BY ROB BECOTTE 30 APRIL 2020

COMPANY ENGAGEMENT ACTIVITY SINCE INCEPTION



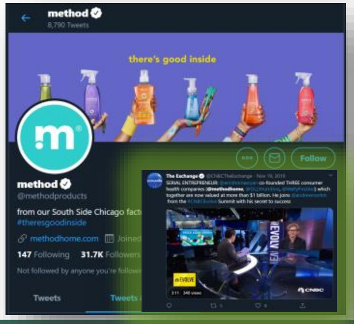
Mrs. Meyer's Clean Day
3,425 Tweets

Mrs. Meyer's Clean Day
@MrsMeyersClean
With scents straight from the garden, we aim to inspire Home Makers and celebrate the beautiful things they make – from crafty DIYs to tasty meals.
mrsmeyers.com Joined July 2009
2,080 Following 11.3K Followers
Not followed by anyone you're following



Aesop
3,927 Tweets

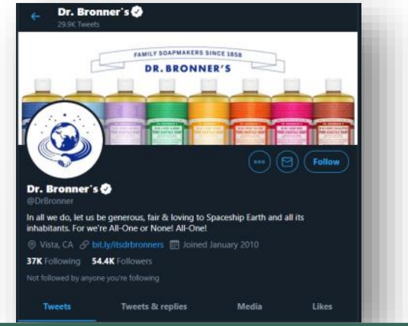
Aesop
@aesopskincare
This is the official Twitter account of Aesop. If seeking skin and hair care counsel, allow us to guide you to @AesopAssistance.
aesop.com Joined December 2011
178 Following 25.1K Followers
Not followed by anyone you're following



method
8,790 Tweets

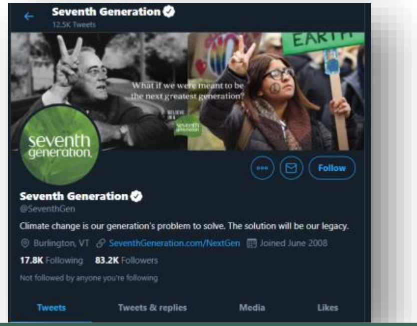
method
@methodproducts
from our South Side Chicago fact
#theresgoodinside
methodhome.com Joined
147 Following 31.7K Followers
Not followed by anyone you're followin

The Exchange @CNBCTheExchange · Nov 19, 2019
SERIAL ENTREPRENEUR: [illegible] co-founded THREE consumer health companies (@methodhome, [illegible], [illegible]) which together are now valued at more than $1 billion. He joins [illegible] from the #CNBCEvolve Summit with his secret to success



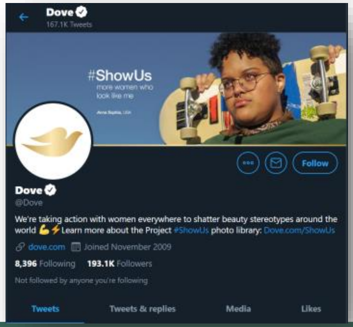
Dr. Bronner's
29.9K Tweets

Dr. Bronner's
@DrBronner
In all we do, let us be generous, fair & loving to Spaceship Earth and all its inhabitants. For we're All-One or None! All-One!
Vista, CA bit.ly/itsdrbronners Joined January 2010
37K Following 54.4K Followers
Not followed by anyone you're following



Seventh Generation
12.5K Tweets

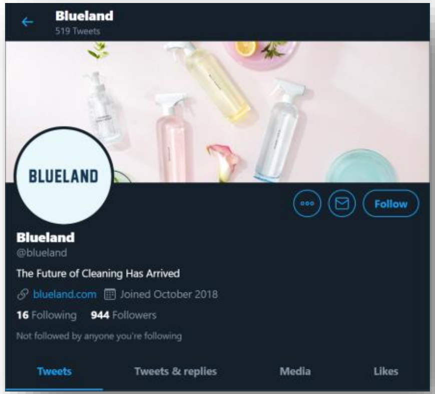
Seventh Generation
@SeventhGen
Climate change is our generation's problem to solve. The solution will be our legacy.
Burlington, VT SeventhGeneration.com/NextGen Joined June 2008
17.8K Following 83.2K Followers
Not followed by anyone you're following



Dove
167.1K Tweets

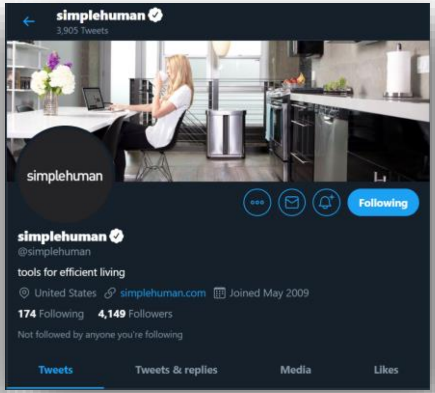
Dove
@Dove
We're taking action with women everywhere to shatter beauty stereotypes around the world 💪⚡Learn more about the Project #ShowUs photo library: Dove.com/ShowUs
dove.com Joined November 2009
8,396 Following 193.1K Followers
Not followed by anyone you're following

[@MrsMeyersClean](#)	[@Aesopskincare](#)	@methodproducts @methodhome [sic]	@DrBronner	@SeventhGen	@Dove
11K followers	25K Followers	32K Followers	54K Followers	83K Followers	193K Followers



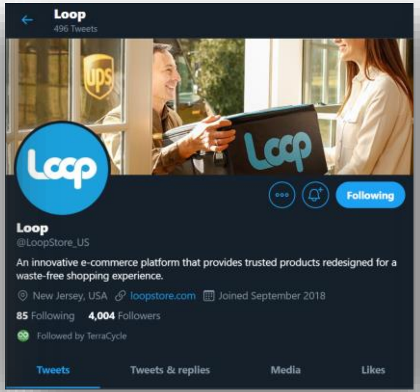
Blueland
519 Tweets

Blueland
@blueland
The Future of Cleaning Has Arrived
blueland.com Joined October 2018
16 Following 944 Followers
Not followed by anyone you're following



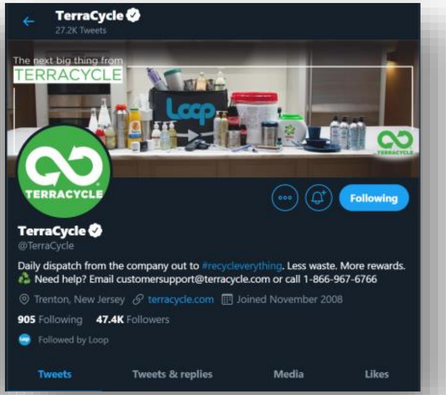
simplehuman
3,905 Tweets

simplehuman
@simplehuman
tools for efficient living
United States simplehuman.com Joined May 2009
174 Following 4,149 Followers
Not followed by anyone you're following



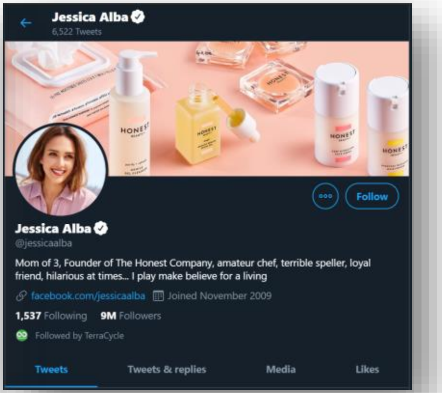
Loop
496 Tweets

Loop
@LoopStore_US
An innovative e-commerce platform that provides trusted products redesigned for a waste-free shopping experience.
New Jersey, USA loopstore.com Joined September 2018
85 Following 4,004 Followers
Followed by TerraCycle



TerraCycle
27.2K Tweets

TerraCycle
@TerraCycle
Daily dispatch from the company out to #recycleverything. Less waste. More rewards. ♻ Need help? Email customersupport@terracycle.com or call 1-866-967-6766
Trenton, New Jersey terracycle.com Joined November 2008
905 Following 47.4K Followers
Followed by Loop



Jessica Alba
6,522 Tweets

Jessica Alba
@jessicaalba
Mom of 3, Founder of The Honest Company, amateur chef, terrible speller, loyal friend, hilarious at times... I play make believe for a living
facebook.com/jessicaalba Joined November 2009
1,537 Following 9M Followers
Followed by TerraCycle

@blueland	@simplehuman	@LoopStore_US	@TerraCycle	@jessicaalba
1K followers	4K Followers	4 K Followers	47K Followers	**9M Followers**

EXECUTIVE SUMMARY

Findings:

1. High positive user-initiated sentiment across examined brands
2. There is evidence validating consumer interest in environmental sustainability #plasticfree
3. @SeventhGen is the consistent leader in sentiment of the three brands on Twitter (Emulate their engagement style), however @LoopStore_US is approaching
4. The Support Vector Machine (SVM) performed 1% better in training data accuracy than the Multinomial Naïve Bayes (MNB) classification algorithm.
5. The MNB algorithm appears to perform better on primary data from Twitter and suggests higher overall brand/consumer sentiment.

CONTENT

1. Orientation
2. Data Collection
3. Model Comparison and Performance
4. Findings
   A. Brand Sentiment
   B. Engagement Time Series Analysis
5. Recommendations and Next Steps

PURPOSE

1. **Catalyst:** Provide an independent data point for a NYC based startup

2. **Hypothesis:** There is high positive consumer sentiment towards using household goods that reduce waste (plastic or one time use packaging) and promote environmental sustainability

3. **End State:** The startup will…

   A. Be armed with current consumer and market atmospherics

   B. Better target branding and marketing efforts

   C. Better align company vision with VCs or existing market leaders (M+A)

YARD STICKS

1. Machine Learning NLP Sentiment Analysis Benchmark for social media (Twitter):

- 60-80% Accuracy Rate

2. Mention Count: A mention is when someone uses the @ sign immediately followed by the Twitter Handle.

- @DrBronner

3. Tag Count: An act of endorsement, which can be very powerful coming from an influencer with an engaged audience made up of people similar to your target market.

- #plasticfree

(This analysis omitted the '#' in the web scrape to capture all data points)

Hypothetical Examples

- **Model accurately predicts positive and negative sentiment in 3 to 4 out of 5 Tweets**
- **"Hey @DrBronner, I love your products!"**
- **"We should live greener #plasticfree**

https://sproutsocial.com/insights/twitter-mentions/

http://nlpprogress.com/english/sentiment_analysis.html

CONTENT

1. Orientation
2. Data Collection
3. Model Comparison and Performance
4. Findings
   A. Brand Sentiment
   B. Engagement Time Series Analysis
5. Recommendations and Next Steps

DATA COLLECTION

Train/Test NLP Data Sets:

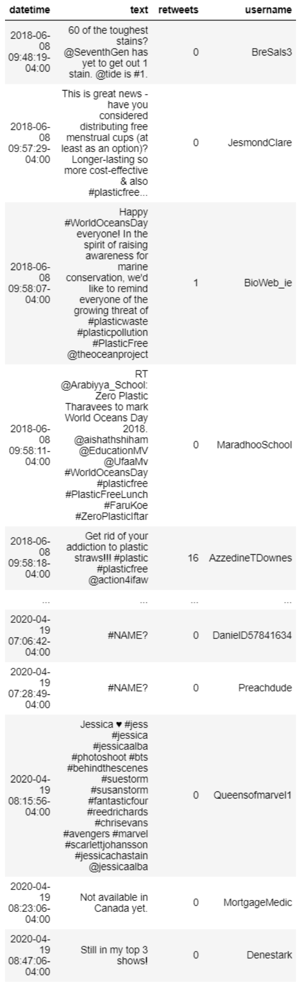
- Kaggle – Twitter and Reddit Tweets (Binary Pos/Neg Labels)
- AWS – 6M Amazon Product Reviews (1-5 Star Label)

Data Scrapes:

- Twitter –GOT3 Python API
  - ~535K non-duplicated Tweets collected from January 1 2018 to April 1 2020
- Reddit – Pushshift Python API
  - ~100K Reddit 'r/SkincareAddiction' posts NOT analyzed, but ready

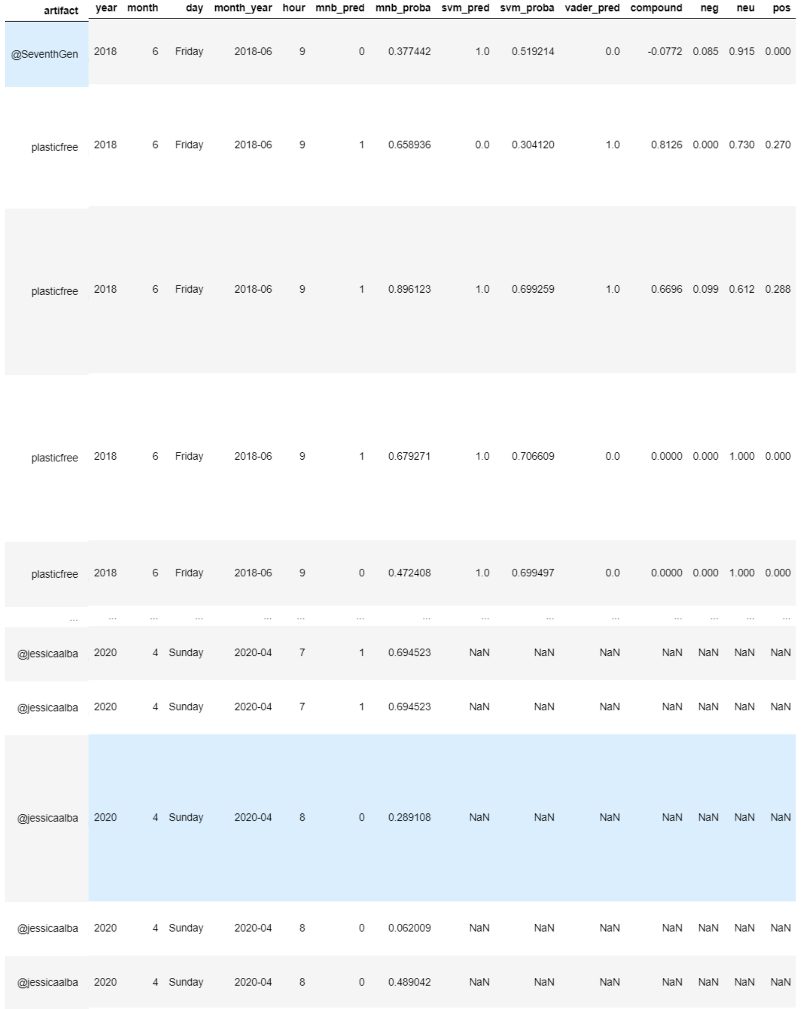
DATA COLLECTION

Scrape Output



datetime	text	retweets	username
2018-06-08 09:48:19-04:00	60 of the toughest stains? @SeventhGen has yet to get out 1 stain. @tide is #1.	0	BreSals3
2018-06-08 09:57:29-04:00	This is great news - have you considered distributing free menstrual cups (at least as an option)? Longer-lasting so more cost-effective & also #plasticfree...	0	JesmondClare
2018-06-08 09:58:07-04:00	Happy #WorldOceansDay everyone! In the spirit of raising awareness for marine conservation, we'd like to remind everyone of the growing threat of #plasticwaste #plasticpollution #PlasticFree @theoceanproject	1	BioWeb_ie
2018-06-08 09:58:11-04:00	RT @Arabiyya_School: Zero Plastic Tharavees to mark World Oceans Day 2018. @aishathshiham @EducationMV @UfaaMv #WorldOceansDay #plasticfree #PlasticFreeLunch #FaruKoe #ZeroPlasticIftar	0	MaradhooSchool
2018-06-08 09:58:18-04:00	Get rid of your addiction to plastic straws!!! #plastic #plasticfree @action4ifaw	16	AzzedineTDownes
...	...	...	...
2020-04-19 07:06:42-04:00	#NAME?	0	DanielD57841634
2020-04-19 07:28:49-04:00	#NAME?	0	Preachdude
2020-04-19 08:15:56-04:00	Jessica ♥ #jess #jessica #jessicaalba #photoshoot #bts #behindthescenes #suestorm #susanstorm #fantasticfour #reedrichards #chrisevans #avengers #marvel #scarlettjohansson #jessicachastain @jessicaalba	0	Queensofmarvel1
2020-04-19 08:23:06-04:00	Not available in Canada yet.	0	MortgageMedic
2020-04-19 08:47:06-04:00	Still in my top 3 shows!	0	Denestark

Feature Engineering (My Additions)



artifact	year	month	day	month_year	hour	mnb_pred	mnb_proba	svm_pred	svm_proba	vader_pred	compound	neg	neu	pos
@SeventhGen	2018	6	Friday	2018-06	9	0	0.377442	1.0	0.519214	0.0	-0.0772	0.085	0.915	0.000
plasticfree	2018	6	Friday	2018-06	9	1	0.658936	0.0	0.304120	1.0	0.8126	0.000	0.730	0.270
plasticfree	2018	6	Friday	2018-06	9	1	0.896123	1.0	0.699259	1.0	0.6696	0.099	0.612	0.288
plasticfree	2018	6	Friday	2018-06	9	1	0.679271	1.0	0.706609	0.0	0.0000	0.000	1.000	0.000
plasticfree	2018	6	Friday	2018-06	9	0	0.472408	1.0	0.699497	0.0	0.0000	0.000	1.000	0.000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
@jessicaalba	2020	4	Sunday	2020-04	7	1	0.694523	NaN	NaN	NaN	NaN	NaN	NaN	NaN
@jessicaalba	2020	4	Sunday	2020-04	7	1	0.694523	NaN	NaN	NaN	NaN	NaN	NaN	NaN
@jessicaalba	2020	4	Sunday	2020-04	8	0	0.289108	NaN	NaN	NaN	NaN	NaN	NaN	NaN
@jessicaalba	2020	4	Sunday	2020-04	8	0	0.062009	NaN	NaN	NaN	NaN	NaN	NaN	NaN
@jessicaalba	2020	4	Sunday	2020-04	8	0	0.489042	NaN	NaN	NaN	NaN	NaN	NaN	NaN

CONTENT

1. Orientation
2. Data Collection
3. Model Comparison and Performance
4. Findings
   A. Brand Sentiment
   B. Engagement Time Series Analysis
5. Recommendations and Next Steps

BASELINE MODEL ACCURACY (OFF THE SHELF ALGORITHM)

VADER Sentiment Analyzer Performance

Data Set	Data Set	Data Set	Data Set
Amazon Reviews	Kaggle Twitter #1 (Indian English Tweets)	Reddit	Kaggle Twitter #2
Long Varied Reviews	Tweet	Posts	Tweet
54% Accuracy	57% Accuracy	63% Accuracy	**64% Accuracy**

CUSTOM MODEL EVALUATION

Model	Tokenize and Fit Computation Time	Best Parameters	Train Accuracy	Test Accuracy
VADER	5 Minutes	OTS	64%	64%
Random Forest	344 Minutes	TFIDF, 20K Tokens Grams: (1,3)	**99%**	76%
Recursive Neural Network	27 Minutes	1 Hidden Layer, 600K Params	78%	76%

Multinomial Naïve Bayes	**2 Minutes**	TFIDF, 40K Tokens,
		n_grams= (1,3),
		use_idf=False
		α= 0.35,
			83%	77%

Support Vector Classifier	360 Minutes	20K Tokens
		C=1.0
		Kernel=‘rbf’
			95%	**78%**

MULTINOMIAL NAÏVE BAYES (CLASSIFIER) TRAINING RESULTS

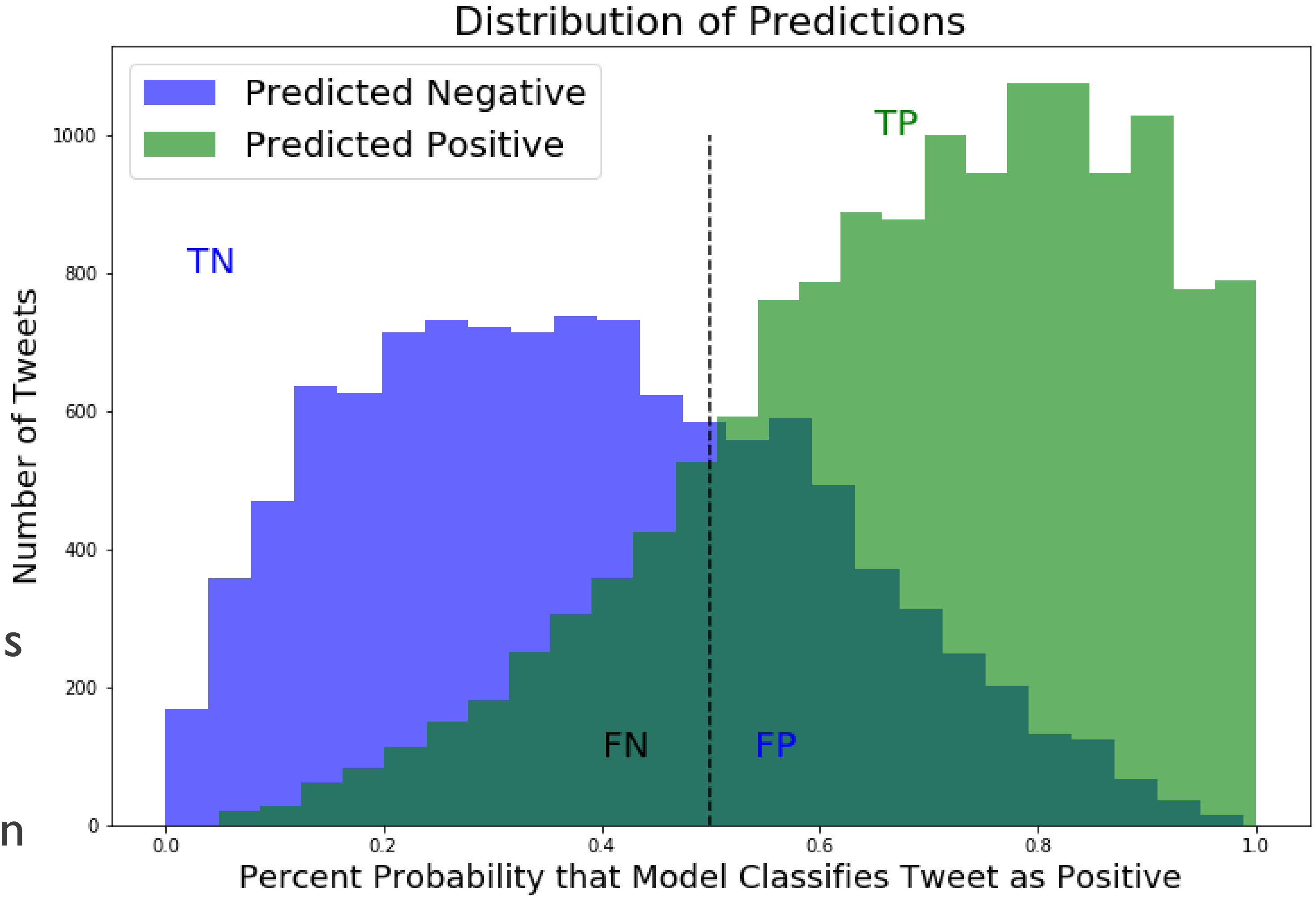
Insight

- Predictions have appropriate skews

  (Most predictions are accurate)

- The high confidence predictions were generally accurate

- This model hedged confidence intervals more than the SVM model, suggests there are many tweets that require more context (Accurate representation of Twitter's often sarcastic tone)



MULTINOMIAL NAÏVE BAYES (CLASSIFIER) TRAINING RESULTS

Insight

- Performs **best** at predicting positive sentiment **(Sensitivity)**

  Business Advice: *Use this model for identifying positive influencers and PR wins*

- **Underperforms** when predicting negative sentiment **(Specificity)**

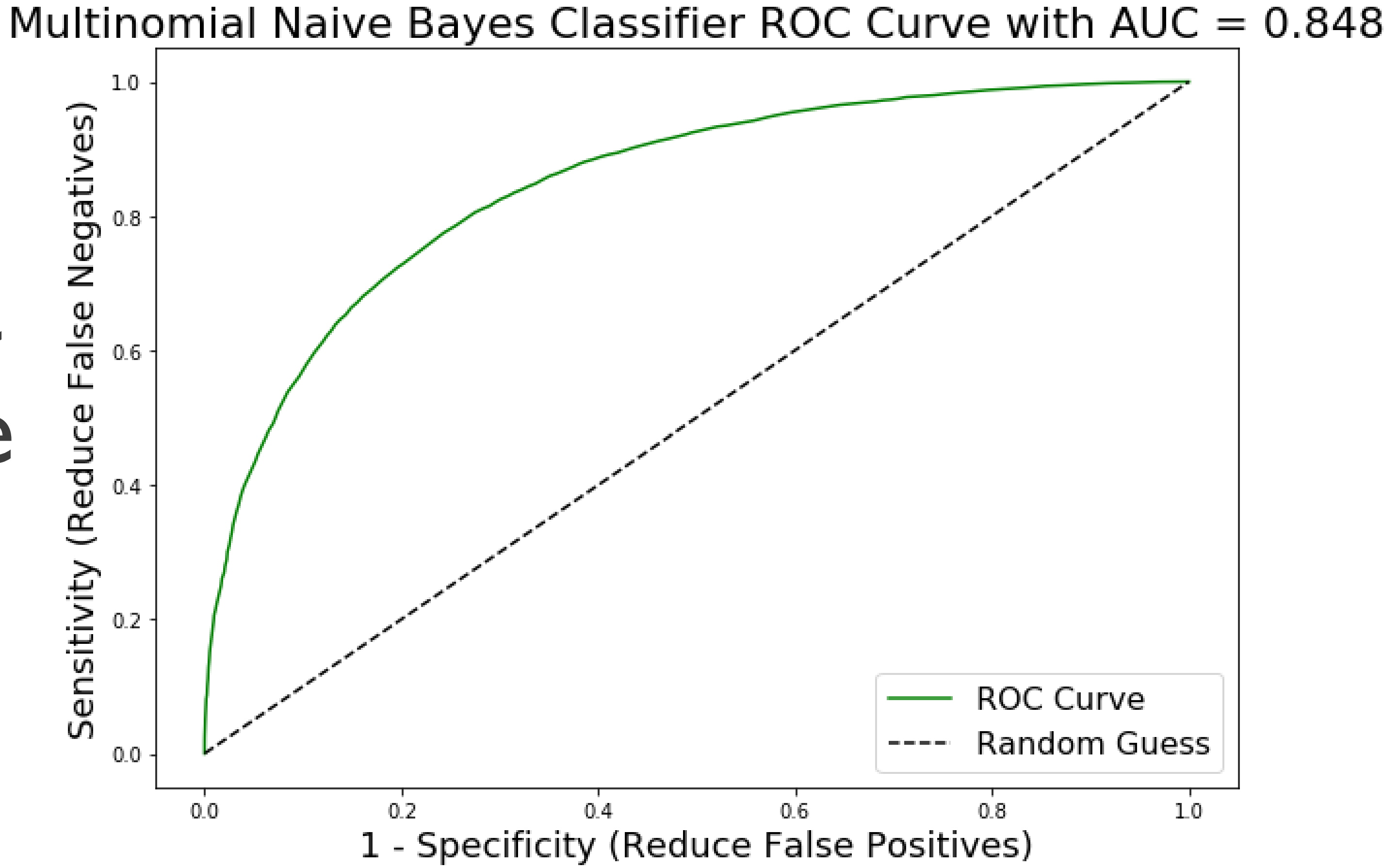
  Business Advice: *Avoid if looking for negative feedback*

	Predicted Negative Tweet	**Predicted Positive Tweet**	
Actual Negative Tweet	7616	Type I Error 3346	Specificity 70%
Actual Positive Tweet	Type II Error 2424	11612	**Sensitivity** **83%**
		Precision 78%	Accuracy 77%

MULTINOMIAL NAÏVE BAYES (CLASSIFIER) TRAINING RESULTS

Insight

- 85% probability of rating a Positive tweet higher than a Negative Tweet

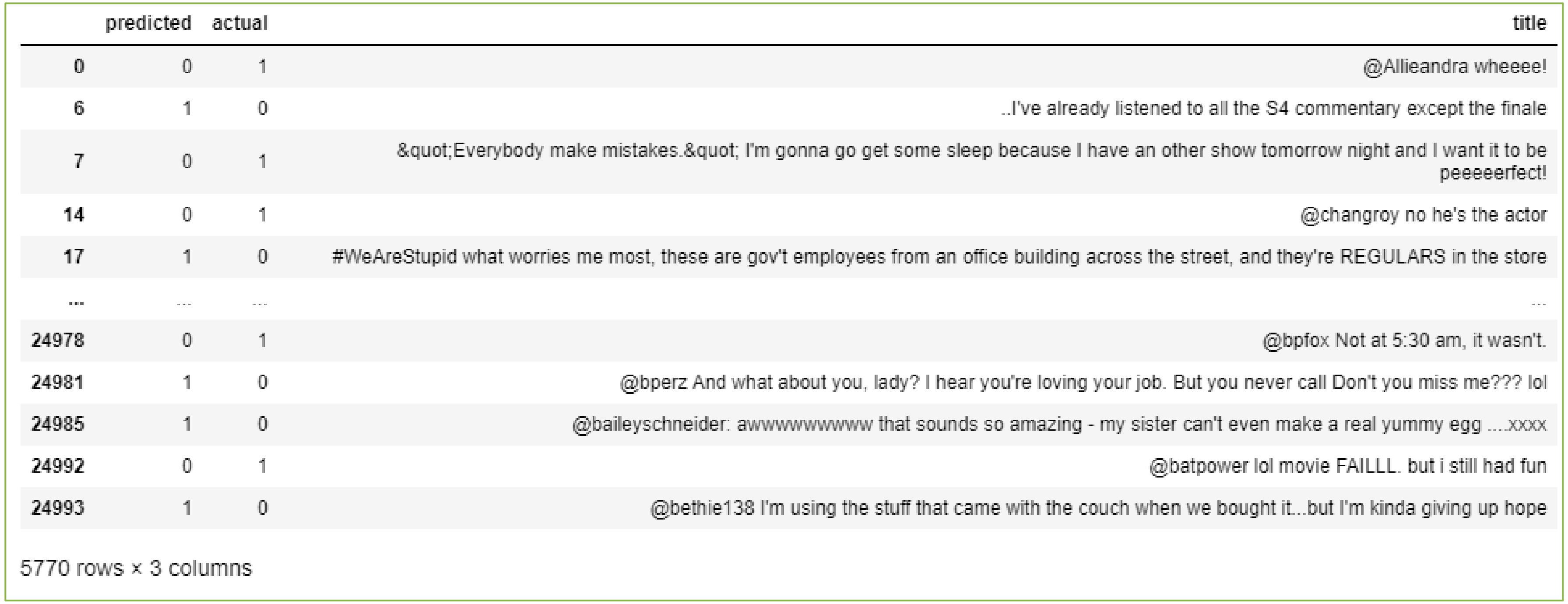


WHERE DID THE MODEL GUESS WRONG?

Twitter Training Data

- Model (and people) need context to make an accurate prediction of sentiment, Ceteris Paribus

- I randomly sampled 10 incorrect guesses… I agreed with the model on 5 of 10



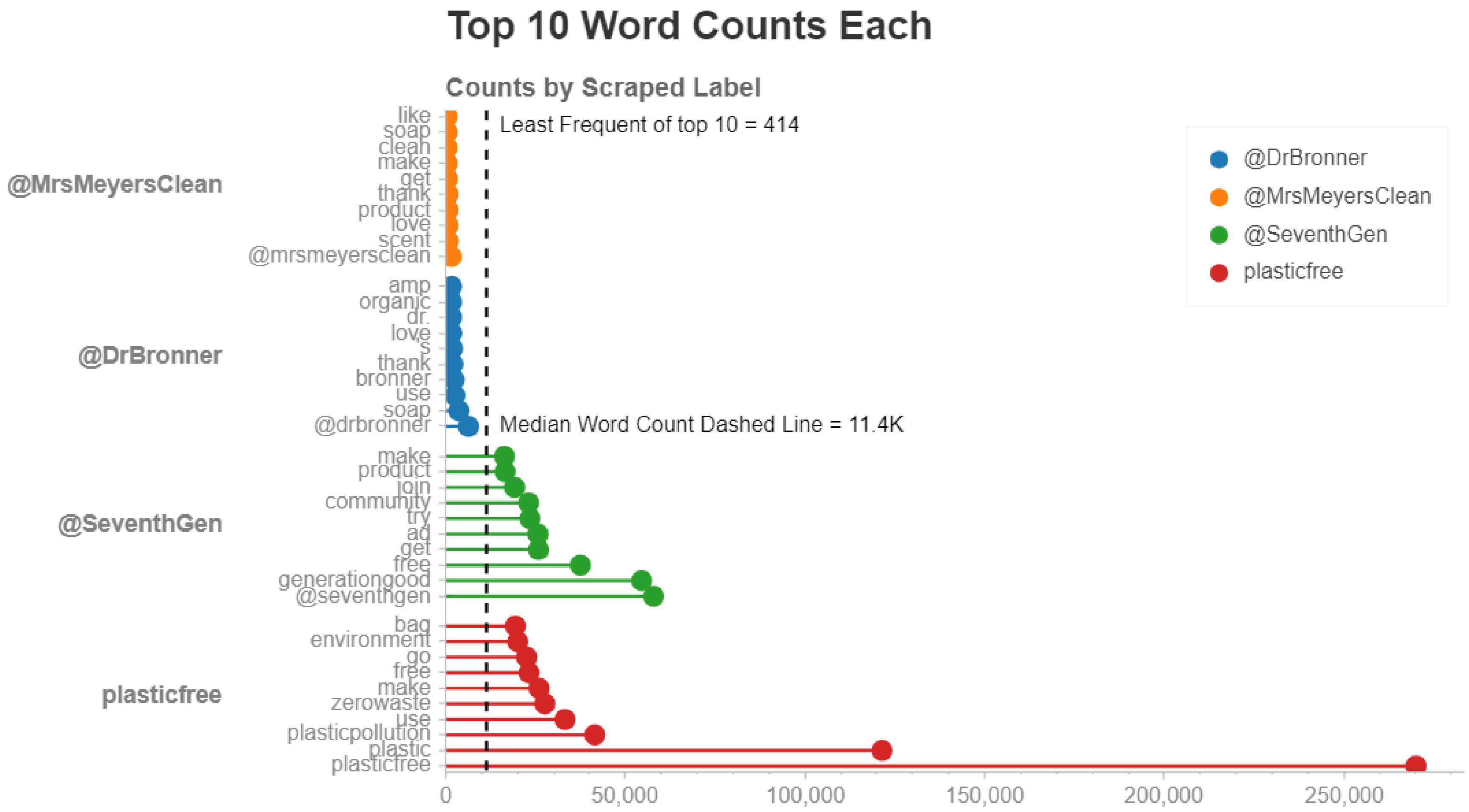
	predicted	actual	title
0	0	1	@Allieandra wheeee!
6	1	0	..I've already listened to all the S4 commentary except the finale
7	0	1	&quot;Everybody make mistakes.&quot; I'm gonna go get some sleep because I have an other show tomorrow night and I want it to be peeeeerfect!
14	0	1	@changroy no he's the actor
17	1	0	#WeAreStupid what worries me most, these are gov't employees from an office building across the street, and they're REGULARS in the store
...	...	...	...
24978	0	1	@bpfox Not at 5:30 am, it wasn't.
24981	1	0	@bperz And what about you, lady? I hear you're loving your job. But you never call Don't you miss me??? lol
24985	1	0	@baileyschneider: awwwwwwwww that sounds so amazing - my sister can't even make a real yummy egg xxxx
24992	0	1	@batpower lol movie FAILLL. but i still had fun
24993	1	0	@bethie138 I'm using the stuff that came with the couch when we bought it...but I'm kinda giving up hope

5770 rows × 3 columns

CONTENT

1. Orientation
2. Data Collection
3. Model Comparison and Performance
4. Findings
   A. Brand Sentiment
   B. Engagement Time Series Analysis
5. Recommendations and Next Steps

TOP 10 MOST OCCURRING WORDS BY INITIALLY EXAMINED BRANDS AND PLASTIC FREE



5.9M tokens in Corpus

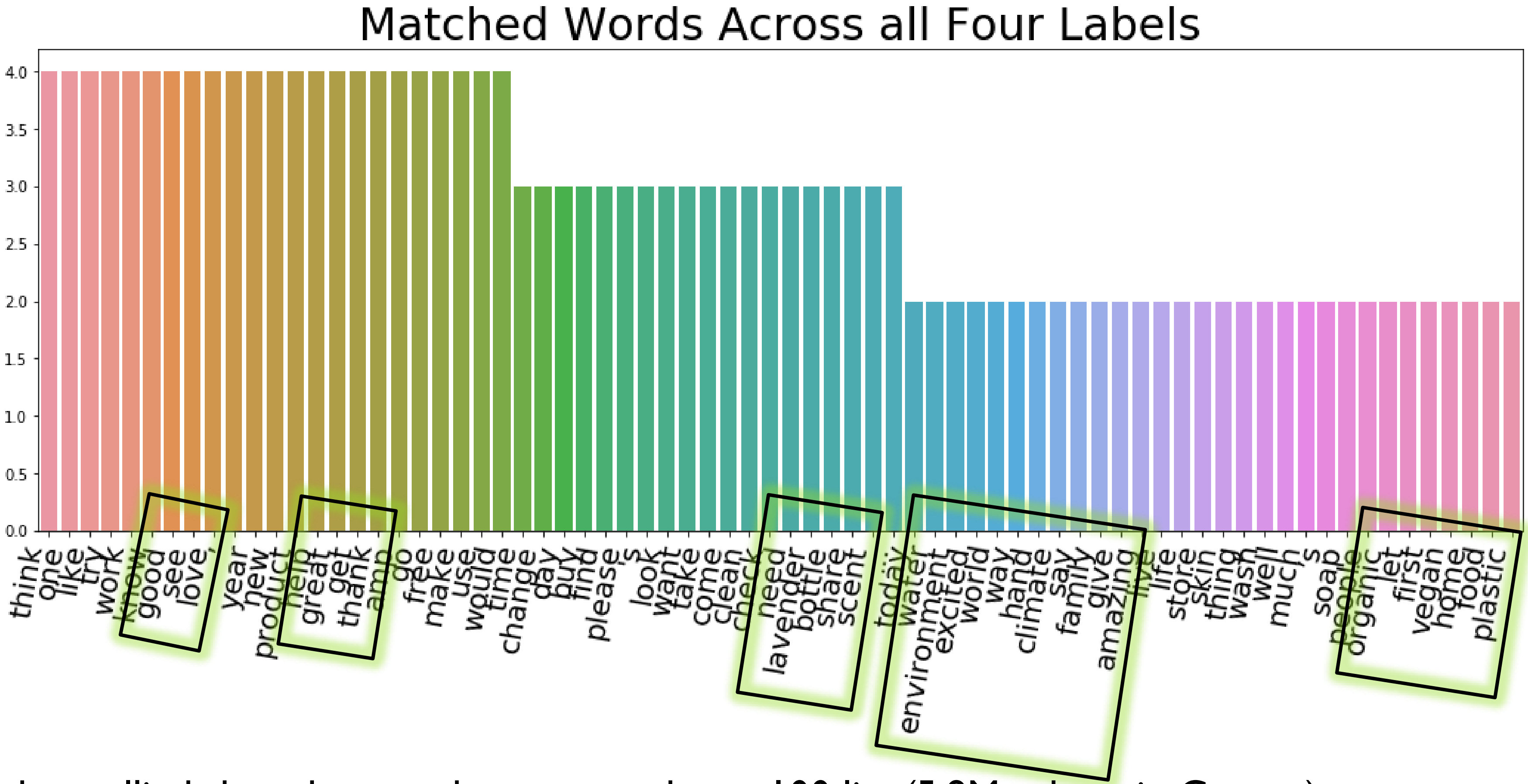
SHARED THEMES BETWEEN ACCOUNTS

Insights

- Highest shared interest in:

1. Positive words
2. Features: ‘Lavender’, ‘scent’, organic’, ‘vegan’, ‘plastic’
3. Environment

Business Advice: *Ensure business philosophy, branding, and actions align with these concepts*



Collated top 100 words from each label, then tallied shared counts between each top 100 list (5.9M tokens in Corpus)

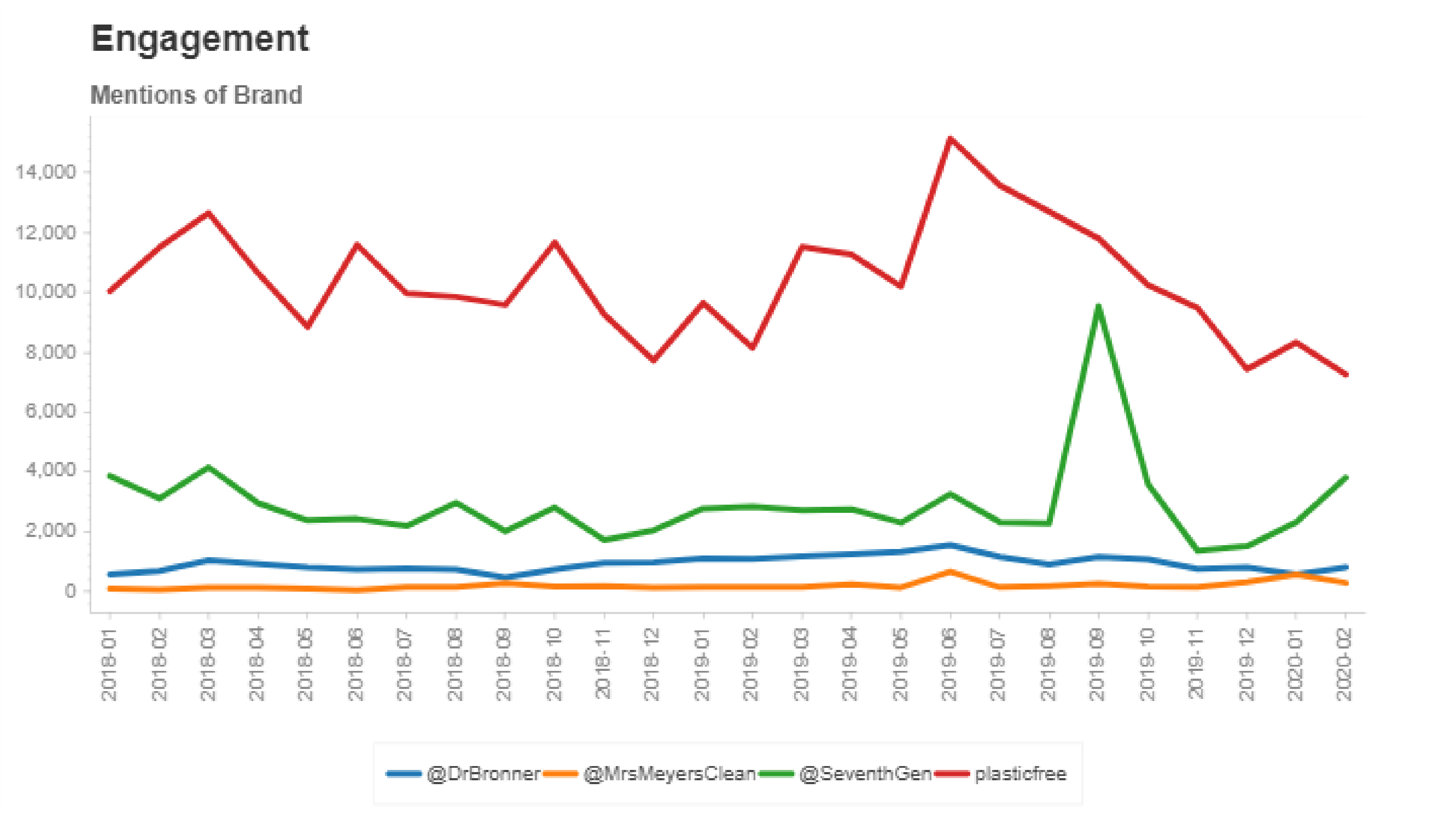
ENGAGEMENT BY MONTH

Insights

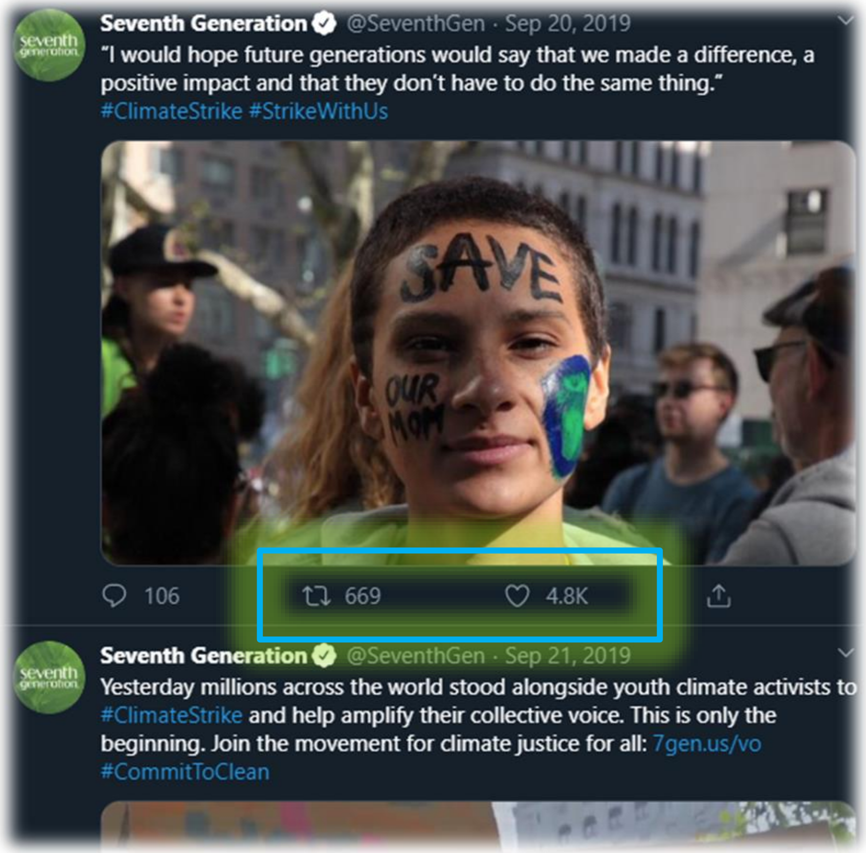
- #plasticfree averages 2.5X the engagement of @SeventhGen

- Generally plateaued mention counts across the board

  Business Advice: *Seek an opportunity to springboard product launch with a high-vis PR event*



#CLIMATESTRIKE



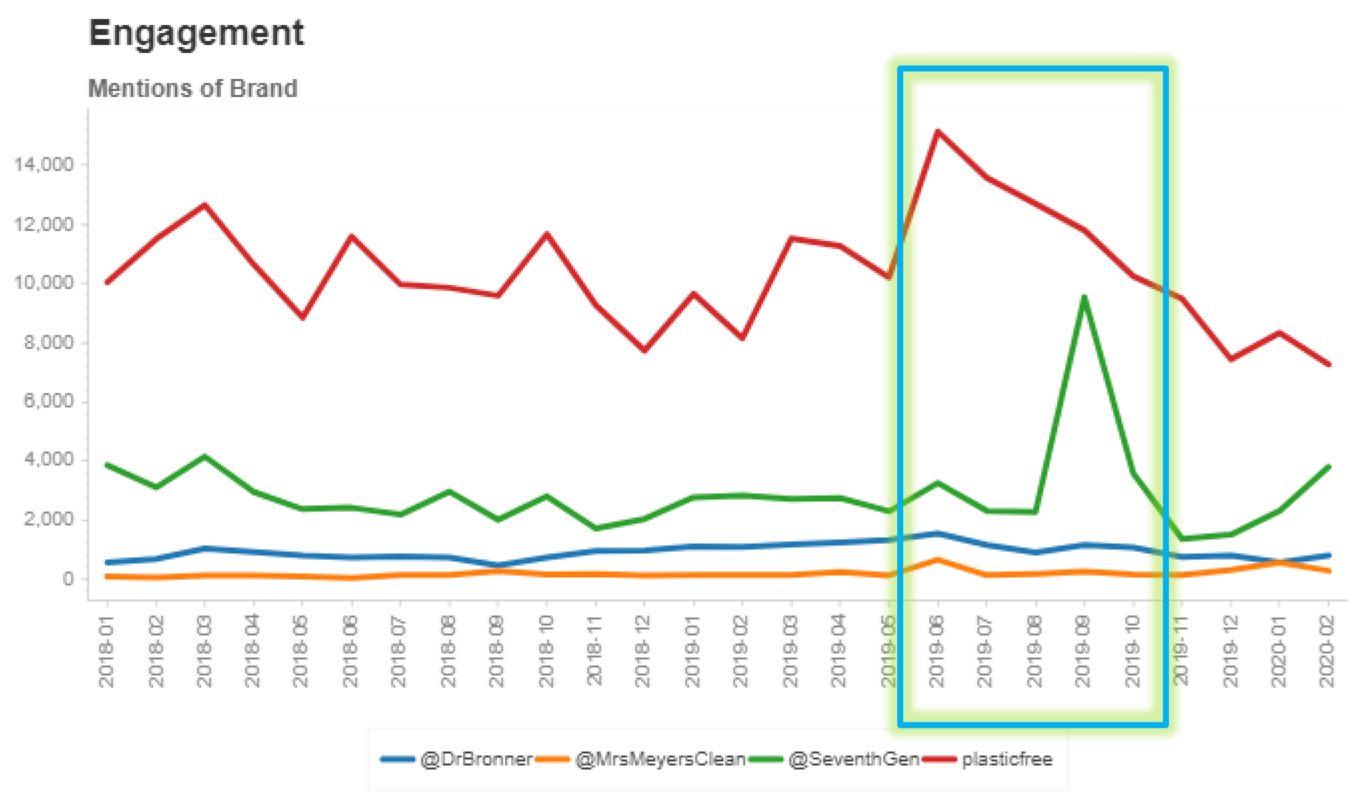
Seventh Generation @SeventhGen · Sep 20, 2019

"I would hope future generations would say that we made a difference, a positive impact and that they don't have to do the same thing."
#ClimateStrike #StrikeWithUs

106 669 4.8K

Seventh Generation @SeventhGen · Sep 21, 2019

Yesterday millions across the world stood alongside youth climate activists to #ClimateStrike and help amplify their collective voice. This is only the beginning. Join the movement for climate justice for all: 7gen.us/vo
#CommitToClean



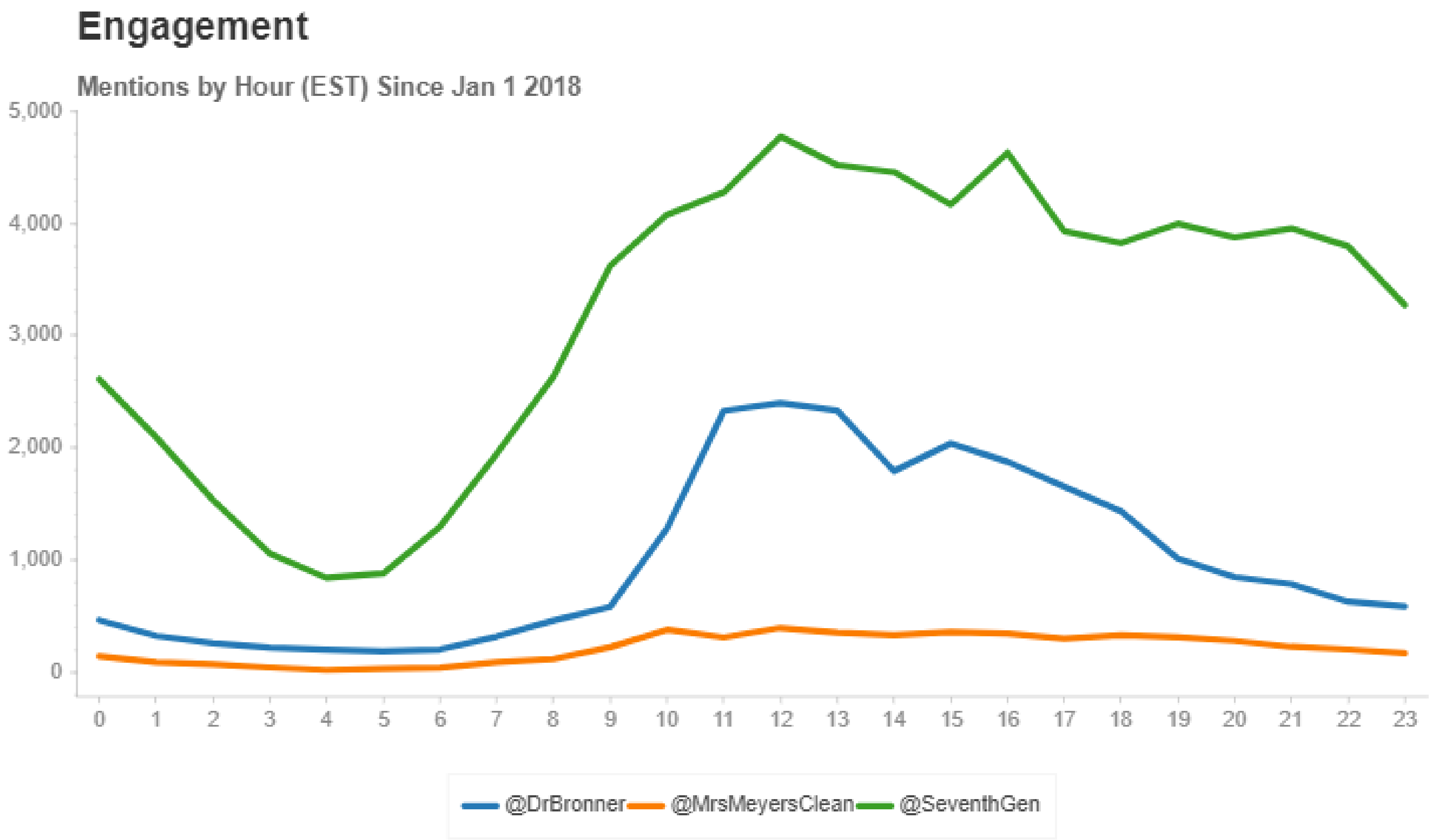
AGGREGATED 2 YEAR ENGAGEMENT BY HOUR

Insights

- Noticeable seasonality:
  - 6AM-12PM Ascent
  - 12AM-5AM Descent

Business Advice:

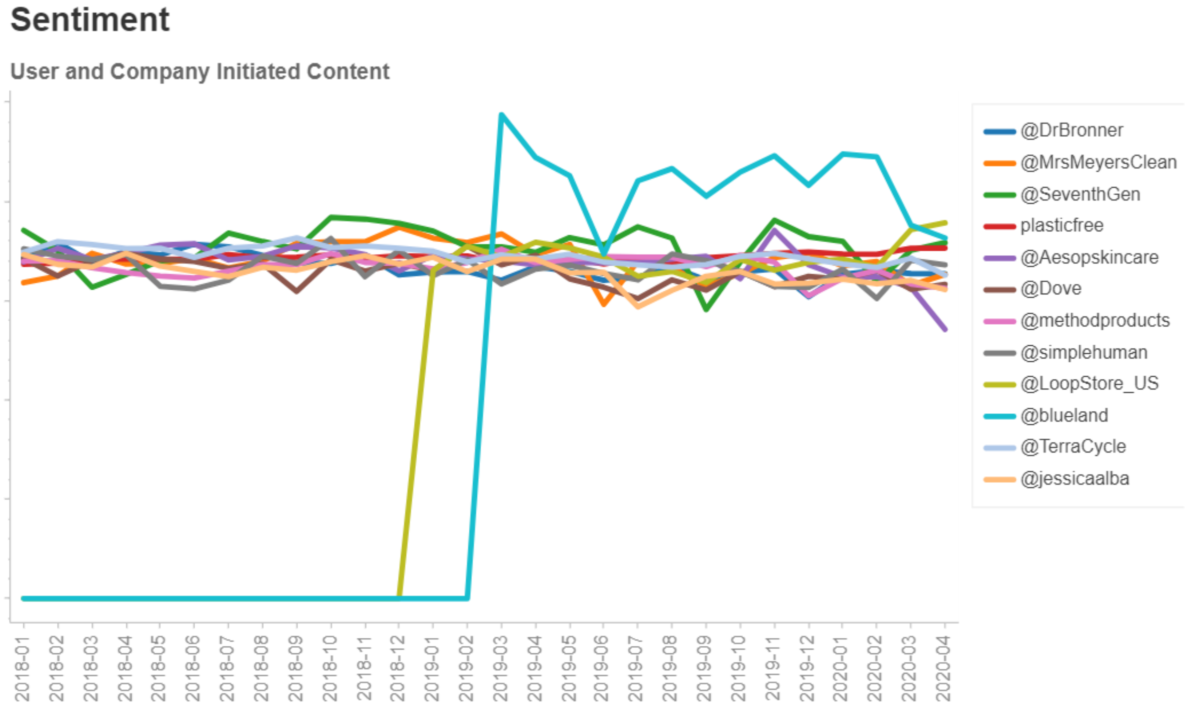
Engage your future consumers when they are active 7AM-7PM



Engagement

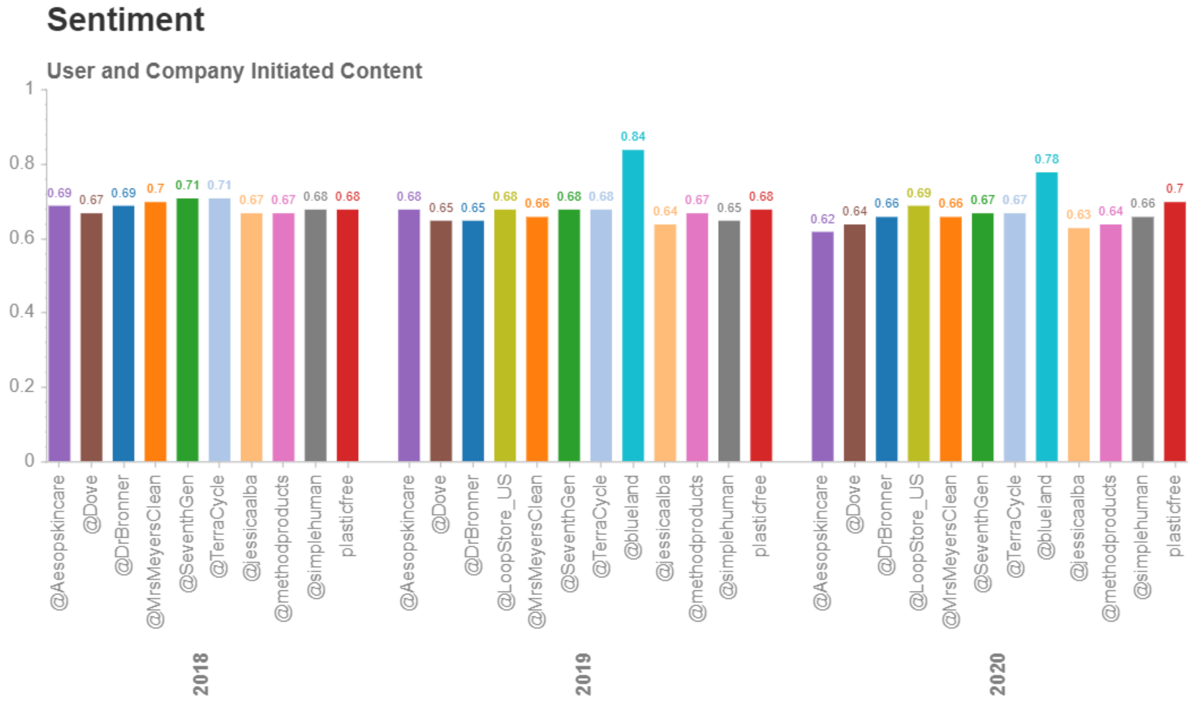
Mentions by Hour (EST) Since Jan 1 2018

USER AND COMPANY GENERATED SENTIMENT



Sentiment

User and Company Initiated Content

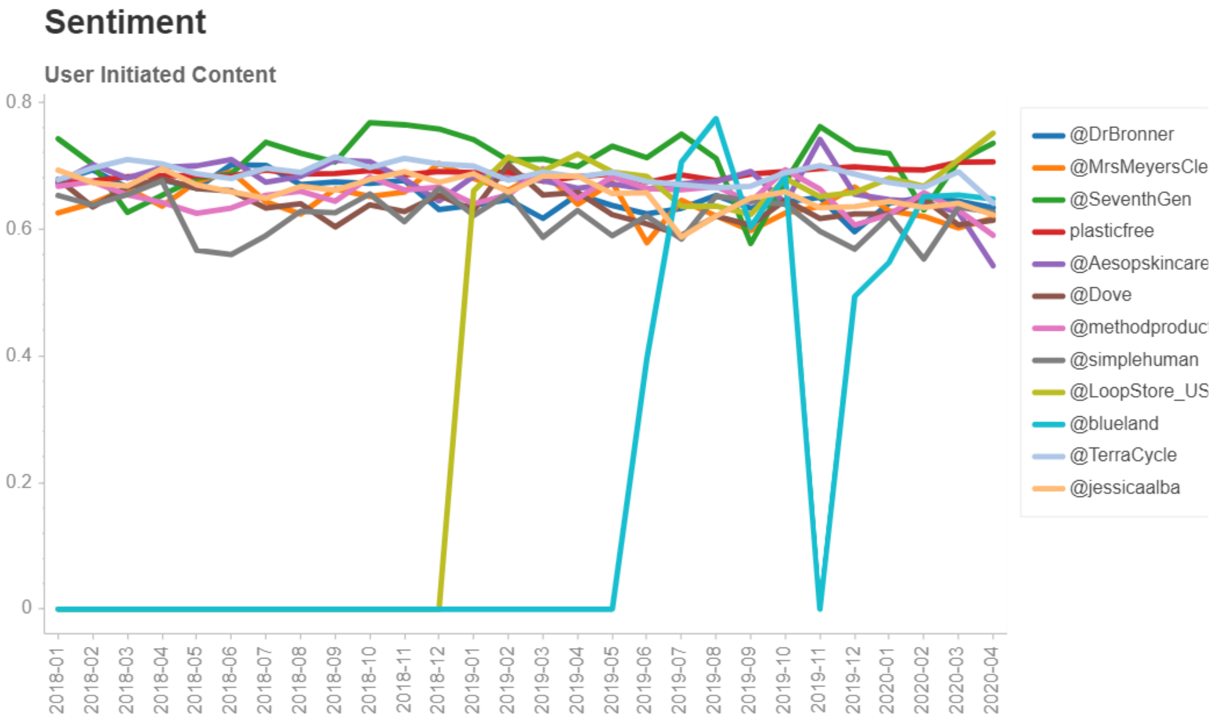


Sentiment

User and Company Initiated Content

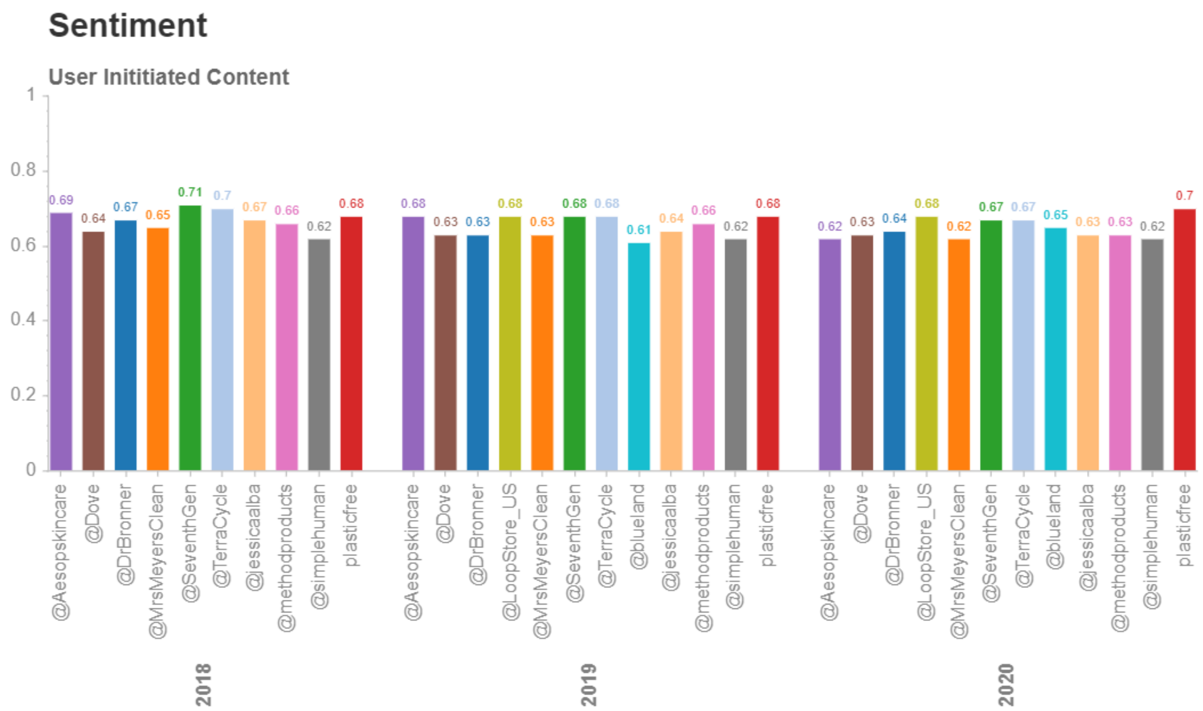
@blueland initially appears to have the highest brand sentiment across the board since inception

USER INITIATED SENTIMENT ONLY



Sentiment

User Initiated Content



Sentiment

User Inititiated Content

- @SeventhGen is the observed leader of consumer sentiment, emulate their engagement style
- Once sentiment is isolated by user-initiated content, @blueland drops significantly
- Takeaway: Companies inherently promote their brand’s image, this over inflates perceived sentiment

CONTENT

1. Orientation
2. Data Collection
3. Model Comparison and Performance
4. Findings
   A. Brand Sentiment
   B. Engagement Time Series Analysis
5. Recommendations and Next Steps

SUMMARY AND NEXT STEPS

Hypothesis:

There is high positive consumer sentiment towards using household goods that reduce waste and promote environmental sustainability

Findings:

1. There is consistent high positive sentiment shared across examined brands (Any score above 50% is classified positive)
2. The data validates high consumer interest in environmental sustainability (#plasticfree is increasingly popular and positive)
3. @SeventhGen is the consistent leader in sentiment of the 10 brands on Twitter (Emulate their engagement style)
4. The Multinomial Naïve Bayes NLP* Classification model performed best on the Sensitivity score (83%)

Recommendations without additional analysis:

1. Align business philosophy, model, branding, and actions consistent with these companies (sustainability)
2. Seek out large scale events and influencers to promote brand and launch
3. Engage consumers while they are active: 7AM to 7PM ET

POTENTIAL NEXT STEPS

Looker deeper at…

1. **Sentiment:**
   - A. Research negative sentiment of competitors
   - B. Research positive sentiment of competitors

2. **Engagement:**
   - A. Time series analysis more granularly
   - B. What are influencers saying
   - C. Cluster analysis
   - D. Industry leader direction signaling

So you can…

- Use findings to address their pitfalls
- Improve on their known successes

- Identify optimal engagement trends to align with your future marketing campaigns
- Look for clusters to identify demographics or unseen driving forces with consumers (Unsupervised ML: Principal Component Analysis)