Augmenting traditional surveillance methods with social media data

April 22nd 2025

Jonah Bregstone

Agenda/Outline

- Social Media Listening Overview
- Project Details
- Implementation
- Visualization & Analysis
- Key Findings / Conclusion

Food Safety Monitoring

- Augmenting the Food and Drug Administration's food and drug safety surveillance systems.
- Traditional surveillance includes hospital records, lab results and official adverse event reporting mechanisms.
- Limitations of conventional methods include:
  - Underreporting
  - Time lag

Social Media Listening - WHY

Social media data provides real-time, organic insights into public health behaviors and opinions, capturing a broader and more immediate range of voices than traditional survey data.

Social Media Listening - Limitations

- Social media data is not fully representative of the general population.
- Social media listening is best used in conjunction with other research activities (e.g., surveys, focus groups)
- Limited user demographic information (e.g., gender, race/ethnicity)
- Social media listening is not a zero-cost or even low-cost research method

Project Overview: Novel Food Products

- RTI worked with a team at the FDA to monitor social media posts about “Novel Food Products”
- To do this the team explored social media data (Twitter/X & Reddit)

Project Overview: Implementation

- Retrieve and store social media data from a variety of sources on key topics & categories.
- Build a custom dashboard to let the research team explore the data.

Accessing Social Media Data

- Direct Retrieval
  - Shifting API landscape
- Commercial Tools
  - These tools have data agreements in place with social media companies
  - Tools used for marketing
  - Can these tools effectively be used by the FDA?

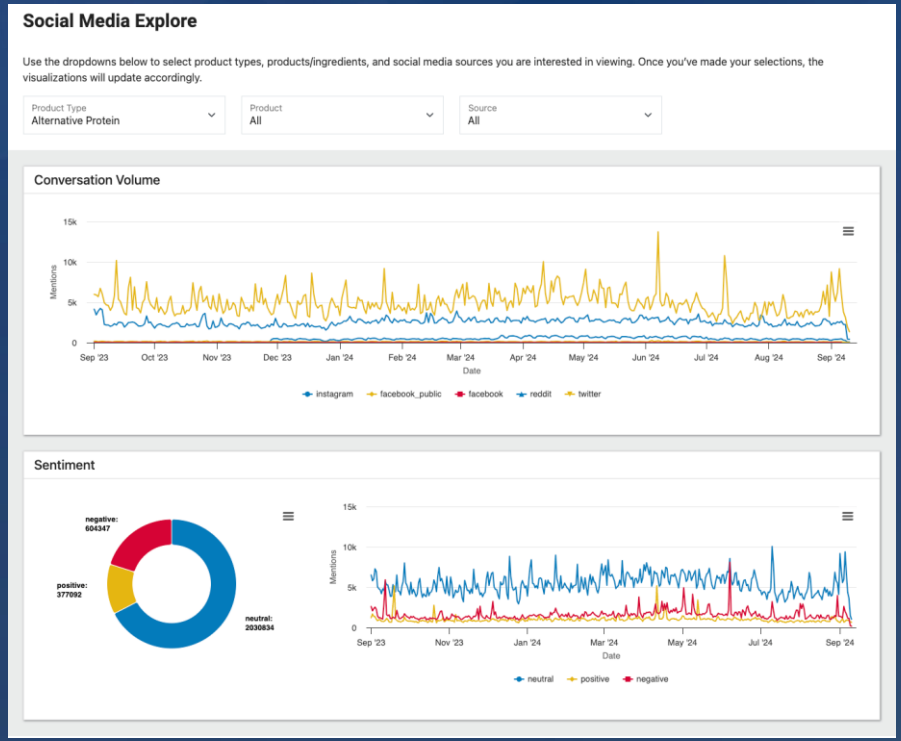
Data Pipeline

- Containerized data-pipeline to retrieve mentions from Brandwatch daily
- Writes to PostgreSQL database on AWS
  - ~2GB in size
  - ~1.5 million mentions
- How do we summarize that data?

x

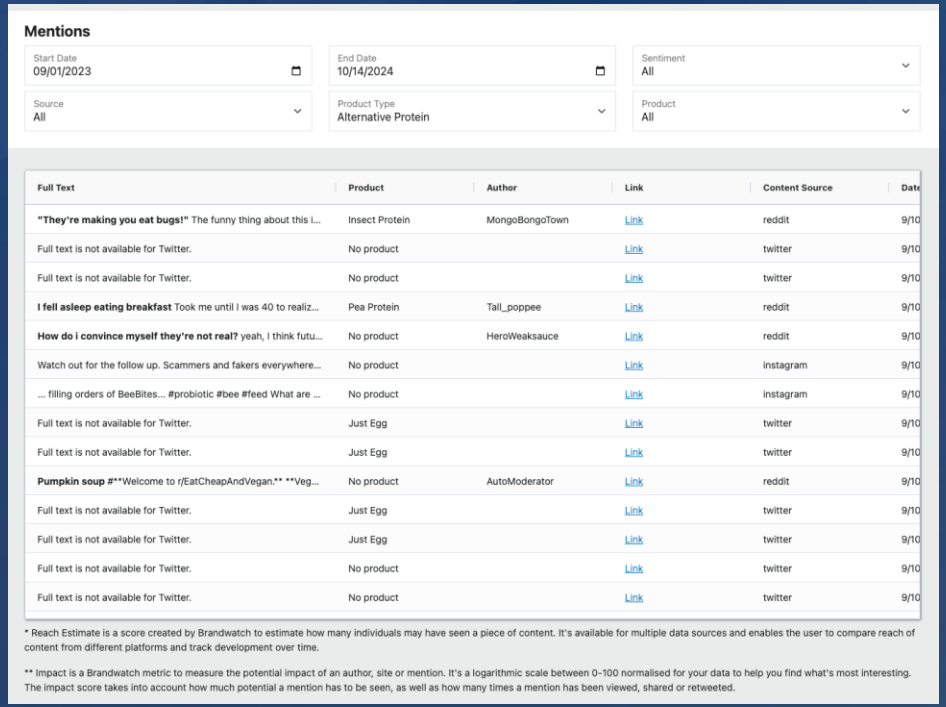
Data Pipeline - Mentions

- Visualize common/ important mention metrics over time



Data Pipeline - Mentions

- Present as tabular data for deeper exploration
- How can we provide insight?



Mentions

Start Date 09/01/2023 | End Date 10/14/2024 | Sentiment All

Source All | Product Type Alternative Protein | Product All

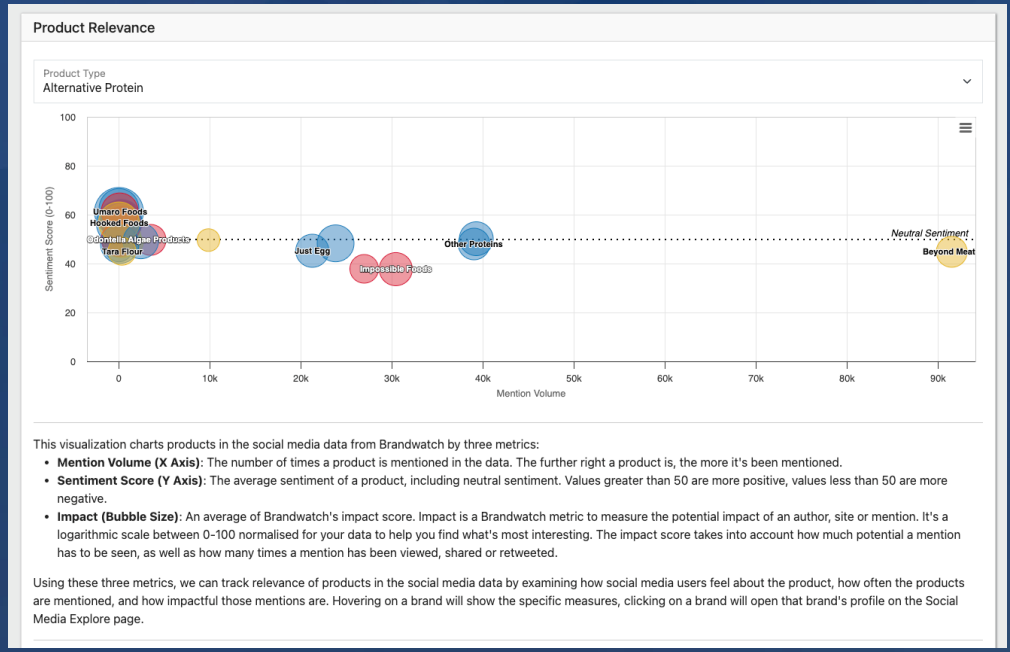
Full Text	Product	Author	Link	Content Source	Date
"They're making you eat bugs!" The funny thing about this i...	Insect Protein	MongoBongoTown	Link	reddit	9/10
Full text is not available for Twitter.	No product		Link	twitter	9/10
Full text is not available for Twitter.	No product		Link	twitter	9/10
I fell asleep eating breakfast Took me until I was 40 to realiz...	Pea Protein	Tall_poppee	Link	reddit	9/10
How do i convince myself they're not real? yeah, I think futu...	No product	HeroWeaksauce	Link	reddit	9/10
Watch out for the follow up. Scammers and fakers everywhere...	No product		Link	instagram	9/10
... filling orders of BeeBites... #probiotic #bee #feed What are ...	No product		Link	instagram	9/10
Full text is not available for Twitter.	Just Egg		Link	twitter	9/10
Full text is not available for Twitter.	Just Egg		Link	twitter	9/10
Pumpkin soup #Welcome to r/EatCheapAndVegan.** **Veg...	No product	AutoModerator	Link	reddit	9/10
Full text is not available for Twitter.	Just Egg		Link	twitter	9/10
Full text is not available for Twitter.	Just Egg		Link	twitter	9/10
Full text is not available for Twitter.	No product		Link	twitter	9/10
Full text is not available for Twitter.	No product		Link	twitter	9/10

* Reach Estimate is a score created by Brandwatch to estimate how many individuals may have seen a piece of content. It's available for multiple data sources and enables the user to compare reach of content from different platforms and track development over time.

** Impact is a Brandwatch metric to measure the potential impact of an author, site or mention. It's a logarithmic scale between 0-100 normalised for your data to help you find what's most interesting. The impact score takes into account how much potential a mention has to be seen, as well as how many times a mention has been viewed, shared or retweeted.

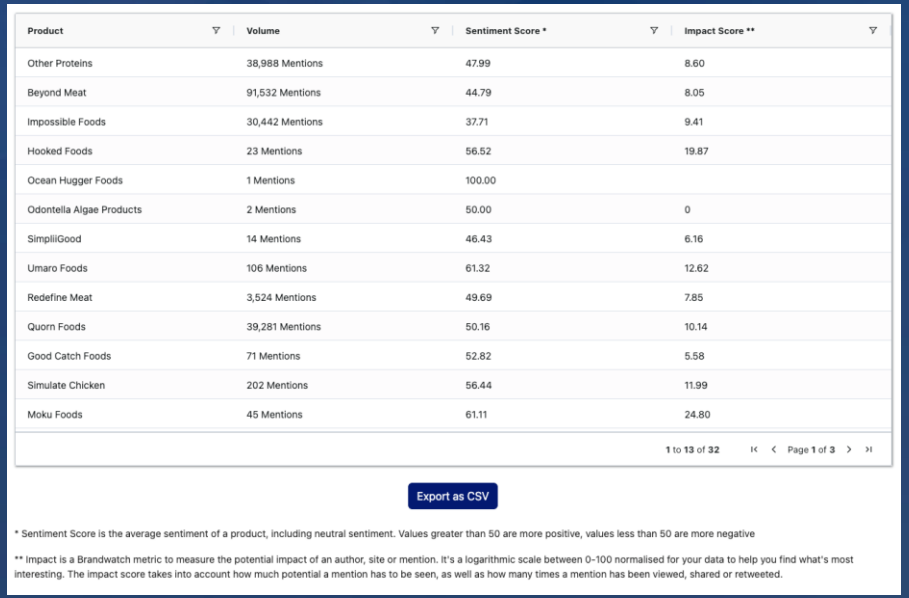
Data Pipeline - Mentions

- Product relevance visualization
- Visual outlier detection
- X: Volume
- Y: Sentiment
- Z: Impact



Data Pipeline - Mentions

- Also available in a tabular format
- Both visual and table are exportable



Product	Volume	Sentiment Score *	Impact Score **
Other Proteins	38,988 Mentions	47.99	8.60
Beyond Meat	91,532 Mentions	44.79	8.05
Impossible Foods	30,442 Mentions	37.71	9.41
Hooked Foods	23 Mentions	56.52	19.87
Ocean Hugger Foods	1 Mentions	100.00	
Odontella Algae Products	2 Mentions	50.00	0
SimpliiGood	14 Mentions	46.43	6.16
Umaro Foods	106 Mentions	61.32	12.62
Redefine Meat	3,524 Mentions	49.69	7.85
Quorn Foods	39,281 Mentions	50.16	10.14
Good Catch Foods	71 Mentions	52.82	5.58
Simulate Chicken	202 Mentions	56.44	11.99
Moku Foods	45 Mentions	61.11	24.80

1 to 13 of 32 Page 1 of 3

Export as CSV

* Sentiment Score is the average sentiment of a product, including neutral sentiment. Values greater than 50 are more positive, values less than 50 are more negative

** Impact is a Brandwatch metric to measure the potential impact of an author, site or mention. It's a logarithmic scale between 0-100 normalised for your data to help you find what's most interesting. The impact score takes into account how much potential a mention has to be seen, as well as how many times a mention has been viewed, shared or retweeted.

Key Findings

- FDA's research interests were well represented in the social media chatter
- Resolvable adverse events were more commonly reported on social media than in official FDA reporting.
- Validated use of social media data in this context – saw adverse events (e.g. Daily Harvest) in the social media data

Lessons Learned

- We also found that by displaying the data as we did, we enabled the team at the FDA to investigate the data and learn about emerging trends
- Increasingly expensive and restricted access to social media data

Thank you

Contact: Jonah Bregstone | email: jbregstone@rti.org