Exploration of Facebook ”Talking About” count

Capstone project for The Data Incubator

Daria Pankova | November 19, 2019

What gets people's attention?

How can one find that out?

Has a marketing event had an effect? Was a controversial incident noticed?

With the data social media collects, answering those questions is possible!

likes, posts, shares, questions, events, photo tags, check-ins, etc.

Facebook records every instance of someone interacting with a company page on Facebook in any impactful way.

Talking about count (TAC)

The number of such interactions by unique users within a week equals to page's "talking about count".

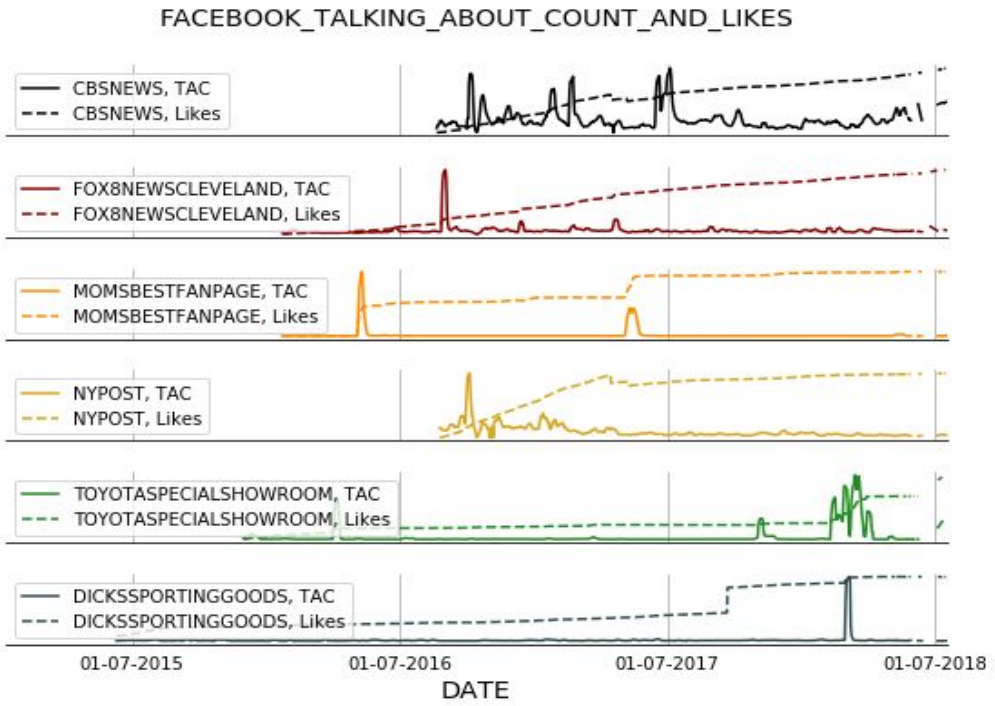
Impactful events

By looking at TAC and sentiment behind it (likes), I would like to create a **dataset** of **"impactful events"** and analyse it.

Dataset: companies' facebook page statistics

I will start by using the dataset Thinknum has made available:

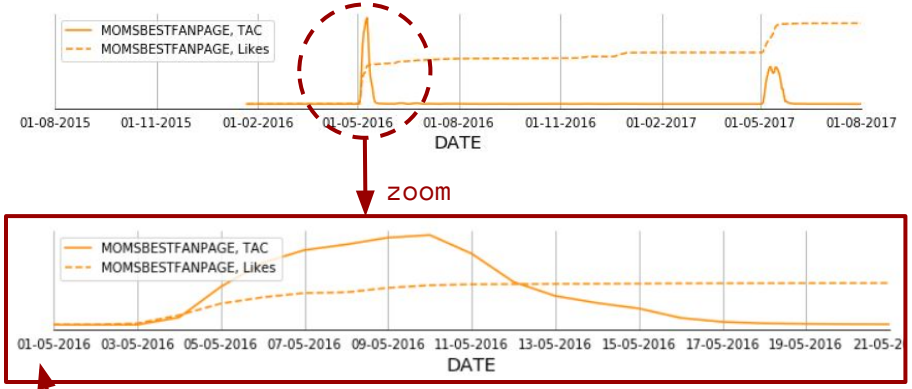
- 432Mb size
- 2015 to mid 2018 (daily update)
- 3.5M entries
- 4950 different companies
- Contains:
  - Number of likes
  - Talking about count
  - Location check-ins



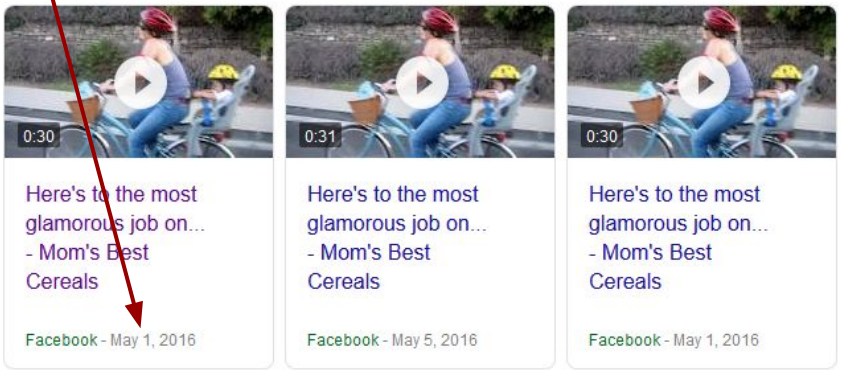
TAC and likes for 6 companies with the biggest spikes in TAC

Confirmation (2 examples)

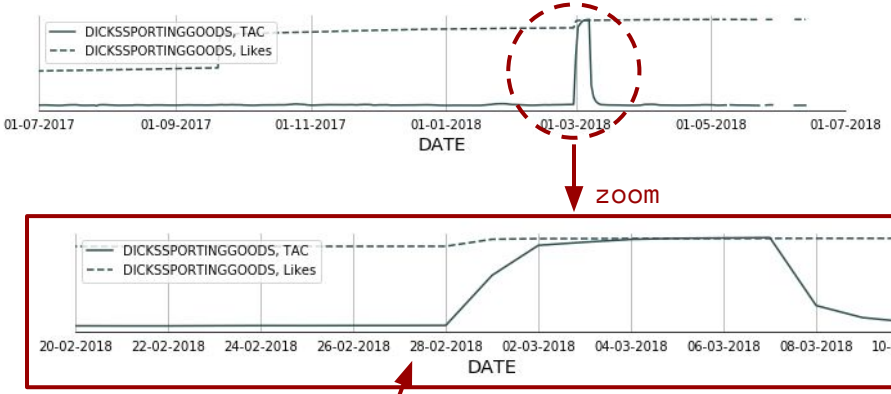
Mom's Best Cereal: **Mother's day video (72M views)**



Videos



Dicks's Sporting Goods: **Weapon sale policy**



MONEY

Dick's Sporting Goods bans sales of assault-style weapons after Parkland, Florida school shooting

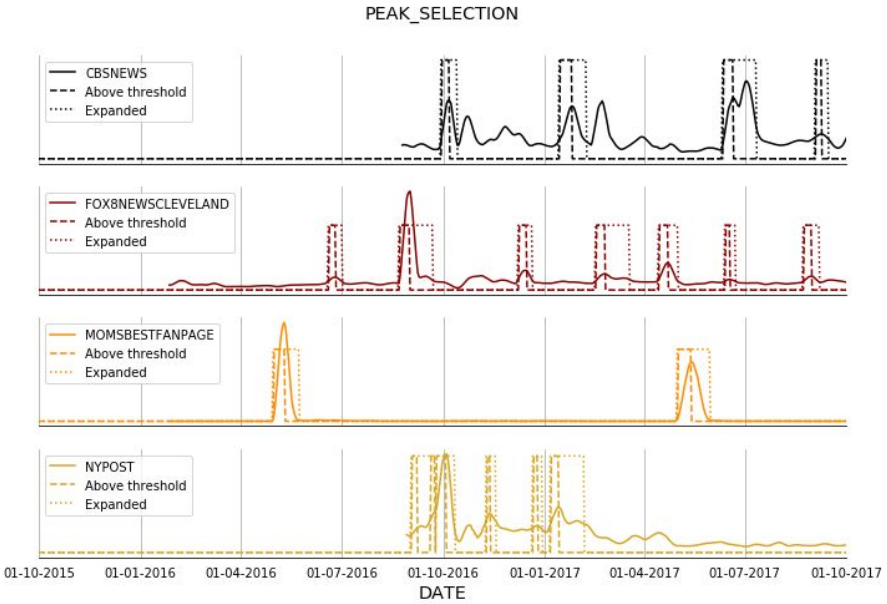
Nathan Bomey USA TODAY

Published 7:31 a.m. ET Feb. 28, 2018 | Updated 12:27 p.m. ET Feb. 28, 2018

Automating peak selection

Complex problem. My best effort so far is shown on the plot:

- **Smooth data**
  - Convolution with a Gaussian
  - Gaussian STD = 2 days
- **Select Peak intervals**
  - Calculate mean and STD for a running interval of 30 day
  - Find times where TAC crosses threshold of 3 STD above the mean
- **Improve duration estimation**
  - Find minimum value for the peak interval
  - Expand interval until the value goes below the minimum of the peak interval



TAC for 4 companies:
Solid line - smoothed TAC,
Dashed line - peak intervals,
Dotted line - expanded intervals.

Dataset (in process)

I record peaks found by algorithm and their characteristics into a dataset.

	name	time	google	facebook	time_window	loc	height_std	height_count	duration	likes
0	CBSNews	[2016-09-27, 2016-09-29]	CBSNews after:2016-09-27 before:2016-09-29	CBSNews after:2016-09-27 before:2016-09-29 sit...	[608, 622]	[608, 622]	6.405724	3.975333e+06	14	0.025468
1	CBSNews	[2017-01-12, 2017-01-14]	CBSNews after:2017-01-12 before:2017-01-14	CBSNews after:2017-01-12 before:2017-01-14 sit...	[715, 739]	[715, 739]	3.780496	2.702325e+06	24	0.043200
2	CBSNews	[2017-06-09,	CBSNews after:2017-06-09	CBSNews after:2017-06-09	[863, 893]	[863,	9.206388	4.505839e+06	30	0.034245

I found ~30 thousand peaks, with current settings.

Next steps:

- ➔ Finding information about events that correspond to the peak.
- ➔ Collecting the information about them automatically or semi-automatically, for statistical analysis. Or focusing on a smaller selection.
- ➔ Interface for the project.

Thank you!