

The background is a teal color with a subtle, repeating pattern of white circuit lines and dots, resembling a printed circuit board. A diagonal white line runs from the top-left corner to the bottom-right corner, dividing the image into two triangular sections.

Data Analysis Findings: BuzzFeed

Why Disinformation Dataset?

- Disinformation is increasing prevalence
- Related to past works in social media context
- What factors resulted in increased article publications?
- Are there any relevant publishers?
- How many times was Donald Trump mentioned?

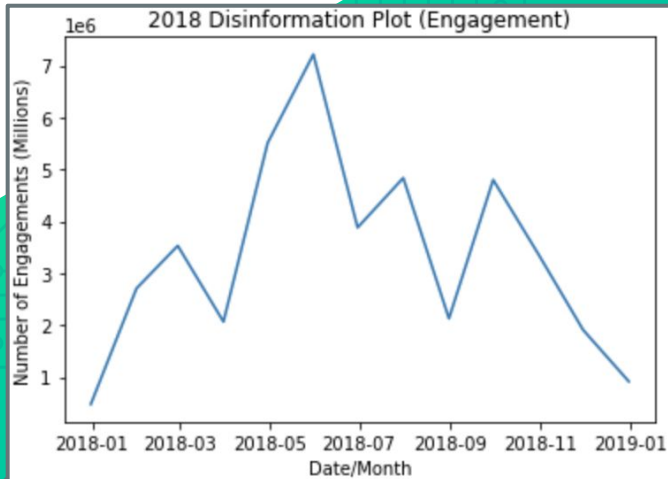
Approach

- Compare timeline of articles published and engagement received
- Find top publishers and categories for more context
 - Many articles did not have categories
- Merged cells of 'category' and 'source' for absent categories required cleaning

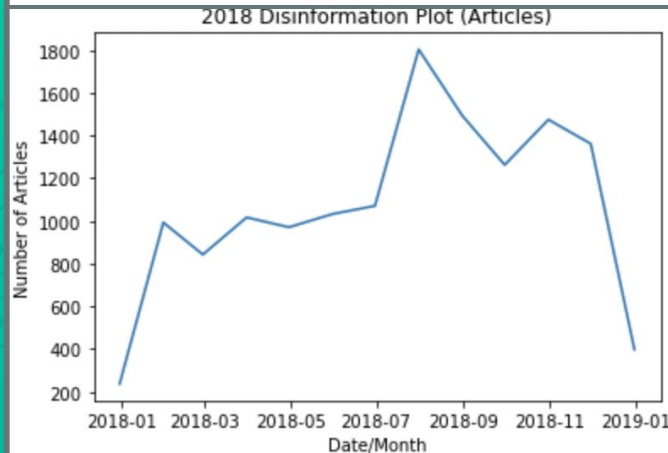
Main Findings

- A spike in article publications can be seen upon a leading month of spiked engagement
 - Otherwise linear increase
- The linear increase in publications can be contributed to political tensions building under Trump-administration

(Retrieved from DataScienceCode.ipynb)



num_engagements	
published_date	
2018-05-31	7215712
2018-04-30	5511081
2018-07-31	4835636
2018-09-30	4798592
2018-06-30	3879637



num_articles	
published_date	
2018-07-31	1804
2018-08-31	1493
2018-10-31	1475
2018-11-30	1362
2018-09-30	1263

Secondary Findings

- Many articles did not include categories due to the spam-nature of disinformation articles
- Many publishers went by an alias rather than name
- Donald Trump was mentioned in 1646/13961 entries (11.7%)

(Retrieved from DataScienceCode.ipynb)

num_articles		num_articles	
source		category	
Sean Adl-tabatabai	2835	NaN	11814
Neon Nettle	1399	Crime	21
Niamh Harris	1284	Politics	10
Baxter Dmitry	1157	Medical	6
Pol Pinoy	654	Business	1

Donald Trump was mentioned in 1646 out of 13961 entries.

Further Project Development: Goals

- Cross multiple datasets to provide increased accuracy and evidence in claims
- Compare multiple years of disinformation
- Under a new administration, is disinformation higher or lower than under the Trump-administration?

Cleaning

- Convert engagement to integers to sum numbers
 - Convert to double, then to integer (no half clicks)
- Strip whitespace and remove emojis from aliases
- Moving publishers to correct column
 - Move merged 'category' & 'source' entries to 'source'
 - Remove categories from 'source' (only publishers)
 - Remove publishers from 'category' (only categories)

Cleaning (Challenging Part)

- Could not remove “everything but” categories in ‘categories’
 - Create a list of unique values in ‘category’
 - Remove the categories from the list
 - For each item in list, remove it in ‘category’