Data & Society Coding Challenge

May 9, 2018

1 Context

We're interested in how YouTube influences are attacking the mainstream media.

One of the most active influencers is Alex Jones, and his media network InfoWars.

In this coding challenge we want

1. The title and statisitcs for videos uploaded by theAlexJonesChannel
2. Filter videos for keywords
3. Analysis of videos about the mainstream media.

2 Problem Set

2.1 Collecting Video Metadata

We want video titles, and viewer staistics for all the videos uploaded to theAlexJonesChannel between between January 1st, 2015 and May 4th, 2018. Save this file as `video_metadata.csv`.

Please take a look at the reference section for the format of the output file, and an example of accessing the YouTube Data API in Python.

Hint: Channel Video IDs/ URLs are available in "user uploads" playlists.

2.2 Filtering the Data

Below are terms our researchers identified as relevant to discourse against Mainstream Media.

```
Mainstream media
MSM
CNN
Mtv
buzzfeed
Nytimes
Ny times
New york times
Wapo
Washington post
Msnbc
Lamestream
propaganda
```

Filter out all videos that do not contain any of the keywords above, and save the output file as `video_metadata_mainstream_media.csv`.

2.3 Analysis of Videos about the Mainstream Media

What is a meaningful insight you can pull from analyzing any of the two datasets you collected?

3 Deliverables

Please send us a zip file containing:

1. CSVs from pt 1 + 2 in a directory called `data`
2. any code/notebooks in a directory called `code`
3. any additional analysis or documents in a directory called `misc`

4 Conclusion

You can work on this assignment using any language and non-Youtube-specific packages. We value D.R.Y., well-organized code that is explainable to non-technical team members. Please spend no more than 1.5 hours on this assignment. Also, submit all deliverables within one week.

5 References

5.1 Youtube API example in Python

Below is an example of accessing the YouTube Data V3 API using Python's Requests library. If you end up using the YT Data API, you'll need to register an API Key on Google Cloud Console.

```
In [ ]: import requests

        key = 'akai-somethingsomething'
        username = 'munchies'

        http_endpoint = ("https://www.googleapis.com/youtube/v3/channels"
                         "?part=id"
                         "&forUsername={0}&key={1}".format(username, key))

        response = requests.get(http_endpoint)
        response_json = response.json()
        print(response_json)
```

5.2 Expected Output

We expect a csv with the following fields (as a minimum!). Please be sure to inlude the video ID as a column named `video_id`.

```
In [6]: output.head()
```

Out[6]:

	video_id	channel_title	channel_id	video_publish_date	video_title	video_view_count	video_like_count
19029	juv2LzTXm28	The Alex Jones Channel	UCvsye7V9psc-APX6wV1twLg	2015-01-01 01:30:01.000000	The Men More Evil Than Kim Jong-un	15718	274.0
19030	tmo9N7X6bYY	The Alex Jones Channel	UCvsye7V9psc-APX6wV1twLg	2015-01-01 01:00:01.000000	Communists Pushing For Violent Revolt	7203	156.0