YouTube Chatter:

Understanding Online Comments Discourse on Misinformative and Political YouTube Videos

Aarash Heydari, Janny Zhang, Shaan Appel, and Xinyi Wu
Professor Gireeja Ranade - UC Berkeley

Motivation:
Compare YouTube Comments

Misinformative

Factual

VS

Political

Apolitical

Data
Media Bias Fact Check

Channel Name	Subscribers	Joined	Factual	Category
Democracy Now	444,805	Jul-05	High	Left Bias
Young Turks	4,224,662	Dec-05	Mixed	Left Bias
Syriana Analysis	27,432	Jan-17	Mixed	Left Bias
New York Times	1,610,431	Oct-06	High	Left-Center Bias
BBC	3,483,096	Apr-06	High	Left-Center Bias
The Guardian	724,629	Feb-06	High	Left-Center Bias
The Dodo	3,046,148	Mar-14	MBFC not found	
Numberphile	2,683,988	Sep-11	MBFC not found	
Science Magazine	119,934	Mar-08	Very High	Pro-Science
PragerU	1,879,540	Jun-09	MBFC not found (political, right-bias)	
RT (Russia Today)	3,234,208	Mar-07	n/a	Questionable Source
InfoWars*	Removed from YouTube		Very Low	Conspiracy-Pseudoscience
Fox	2,186,028	Sep-06	Mixed	Right Bias
Breitbart News	79,288	Feb-15	n/a	Questionable Source
American Renaissance	102,486	Oct-11	n/a	Questionable Source

Data
Media Bias Fact Check

Data Collection
YouTube API

Each channel:

- Up to 200 most recent videos more than two weeks old

Each video:

- Video Metadata
- Direct comments
- Replies
- Comments metadata

Access datasets

Access code for comments-scraping and analysis

Results

Misinformative

Political

VS

Factual

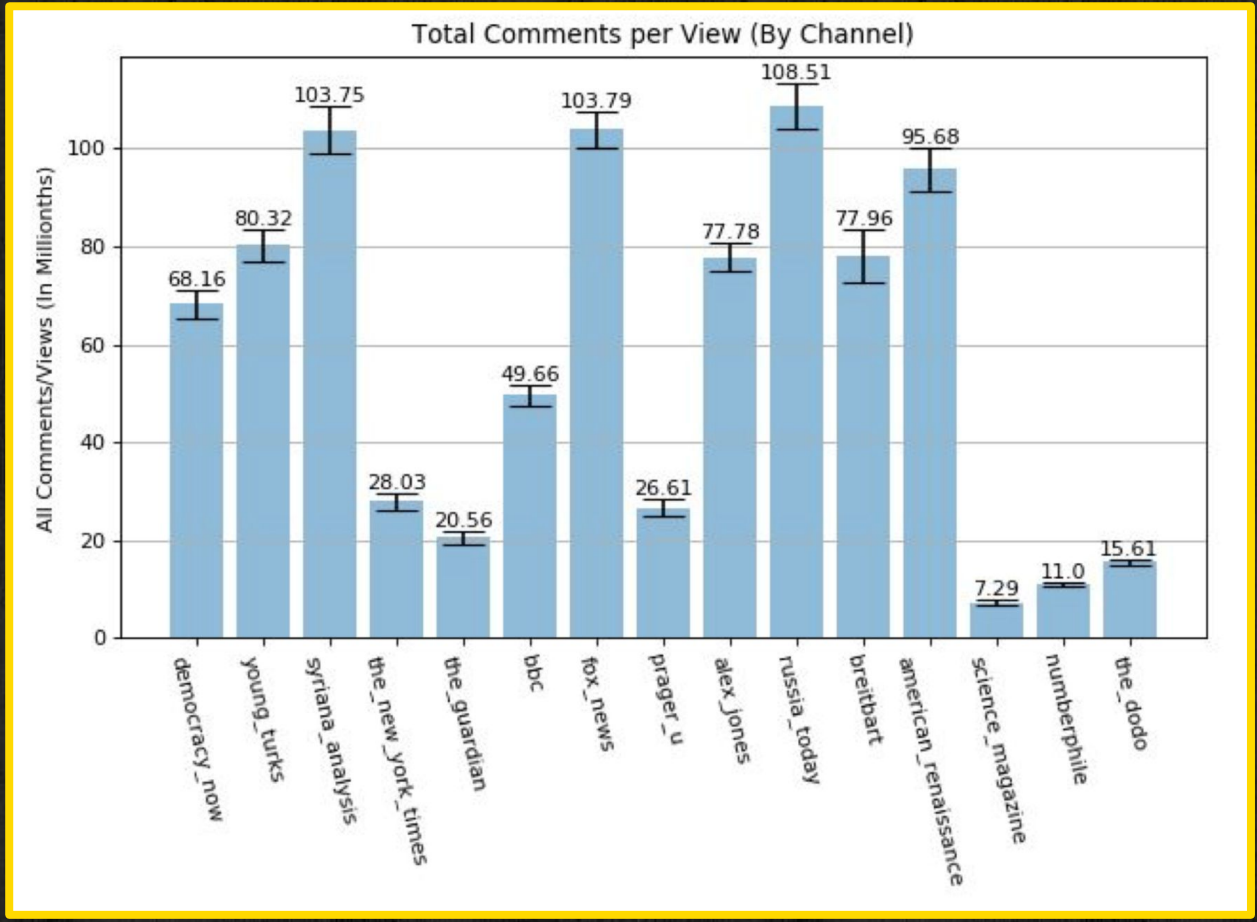
Apolitical

7.5x more comments per view

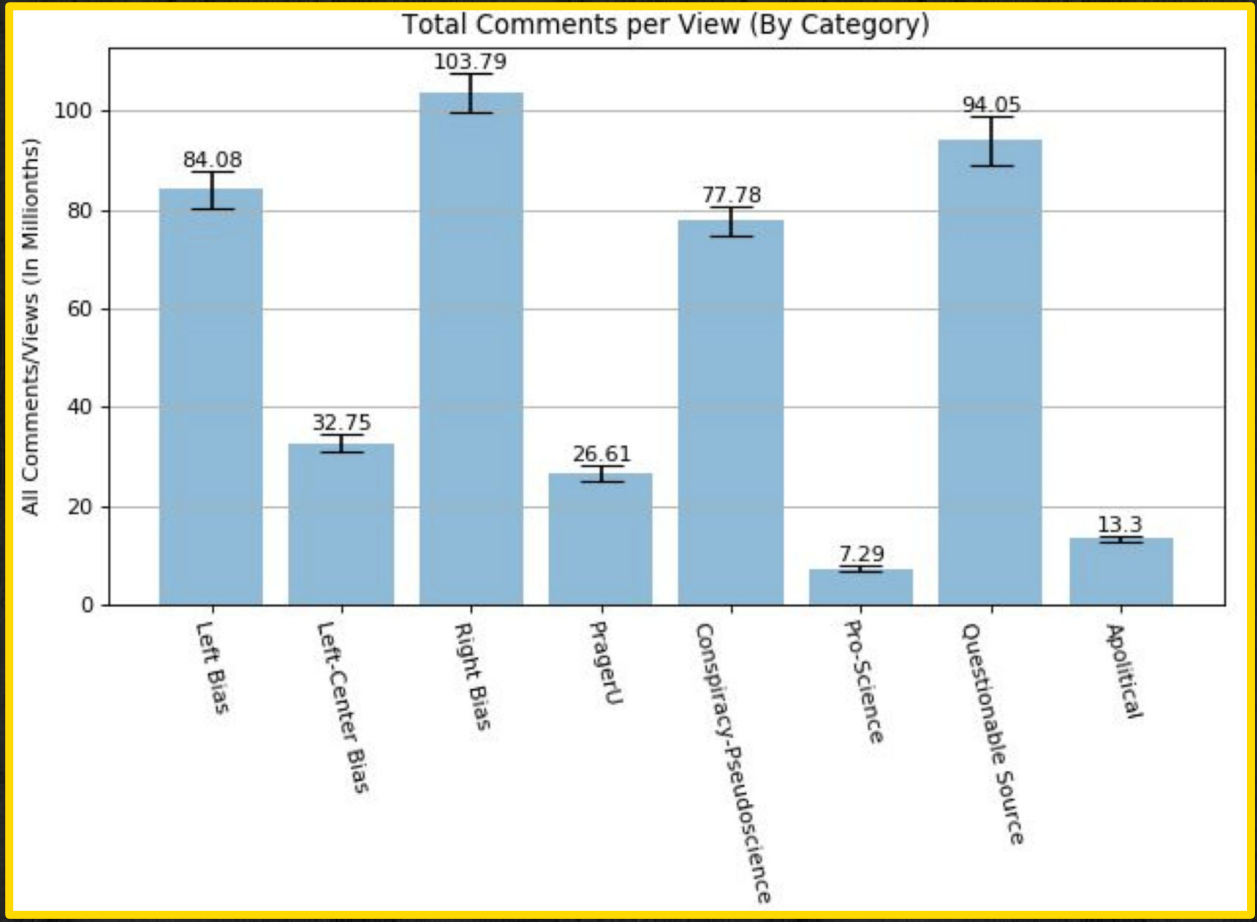
10.42x more replies

More profane comments

Results
Comments per Video by Channel

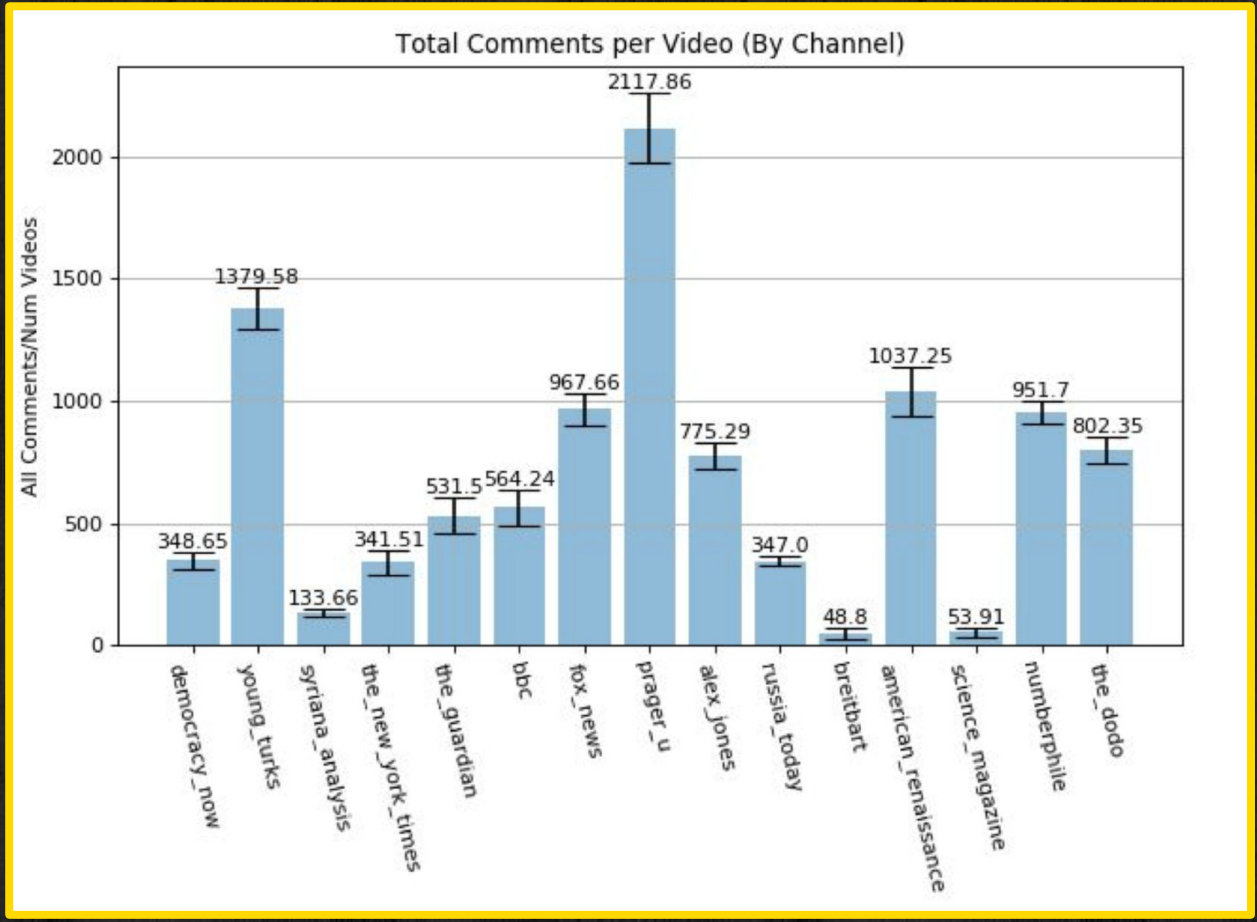


Results
Comments per Video by Category



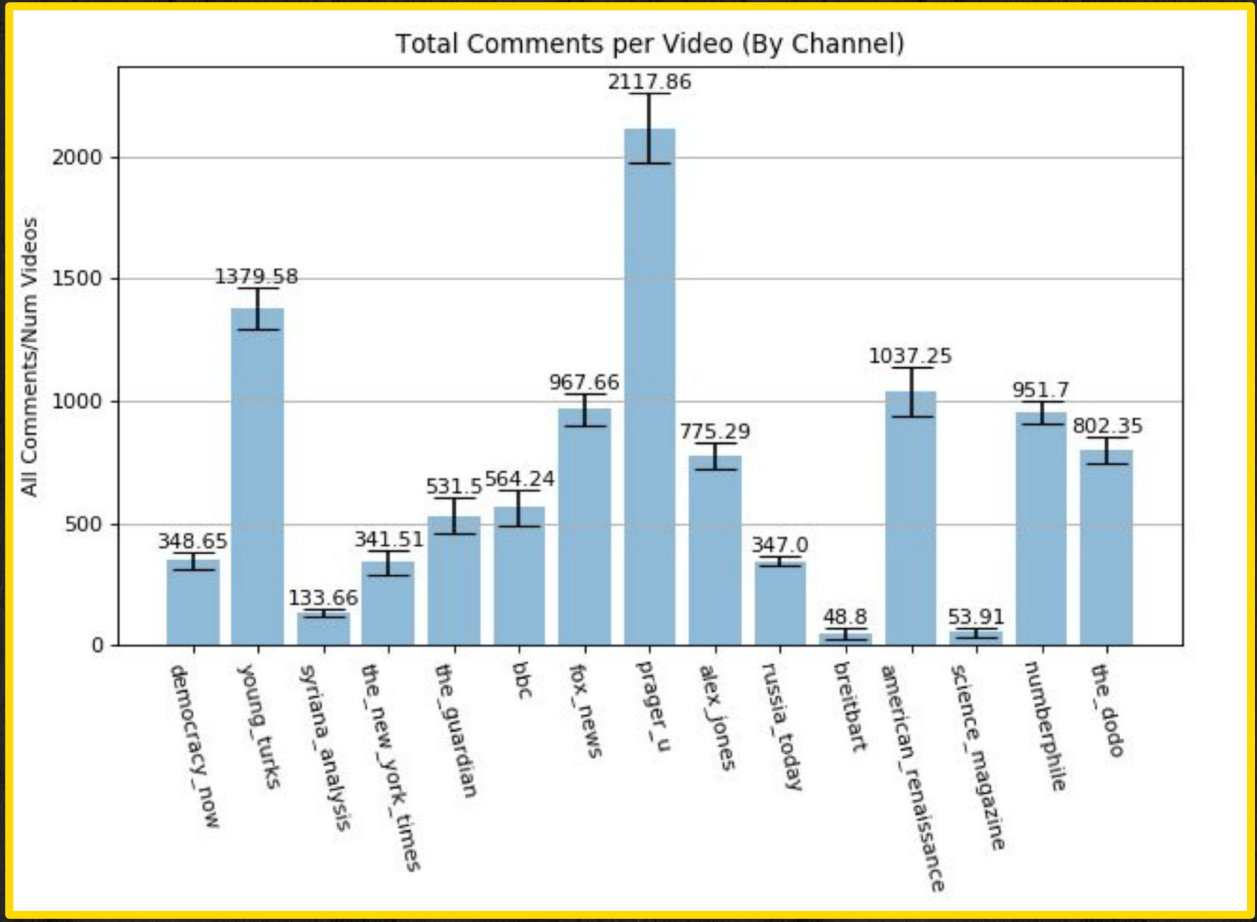
Results

Comments per Video by Channel



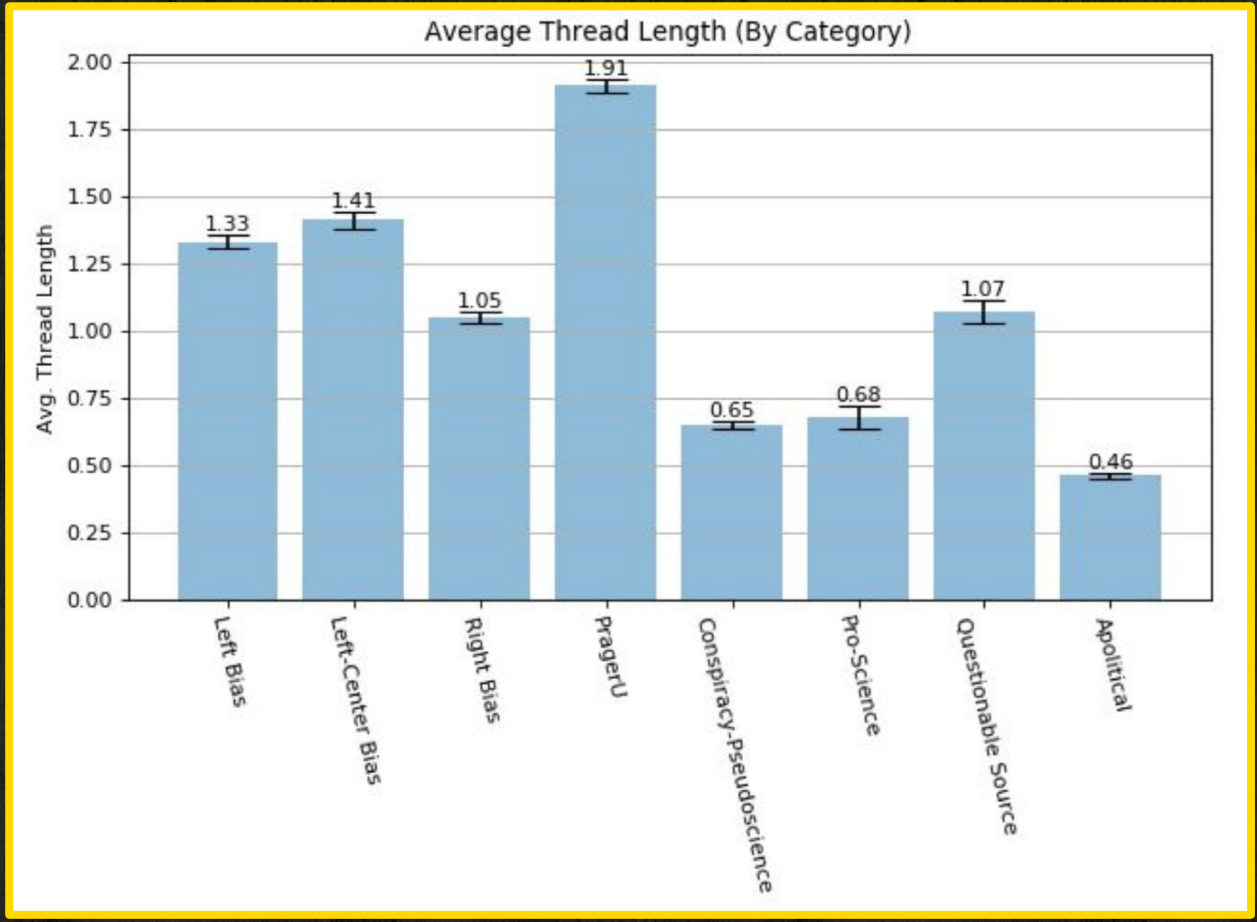
Results

Average Thread Length by Channel



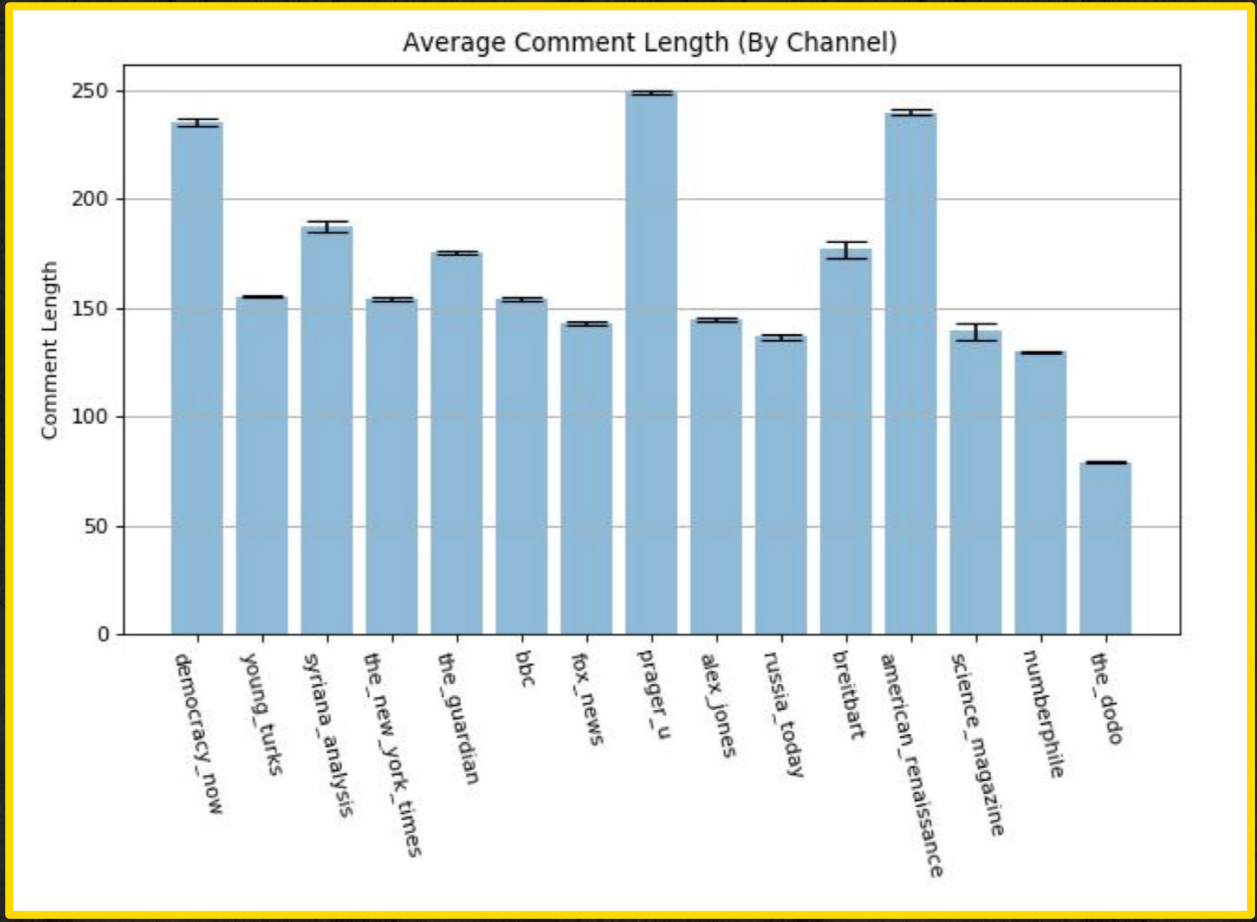
Results

Average Thread Length by Category



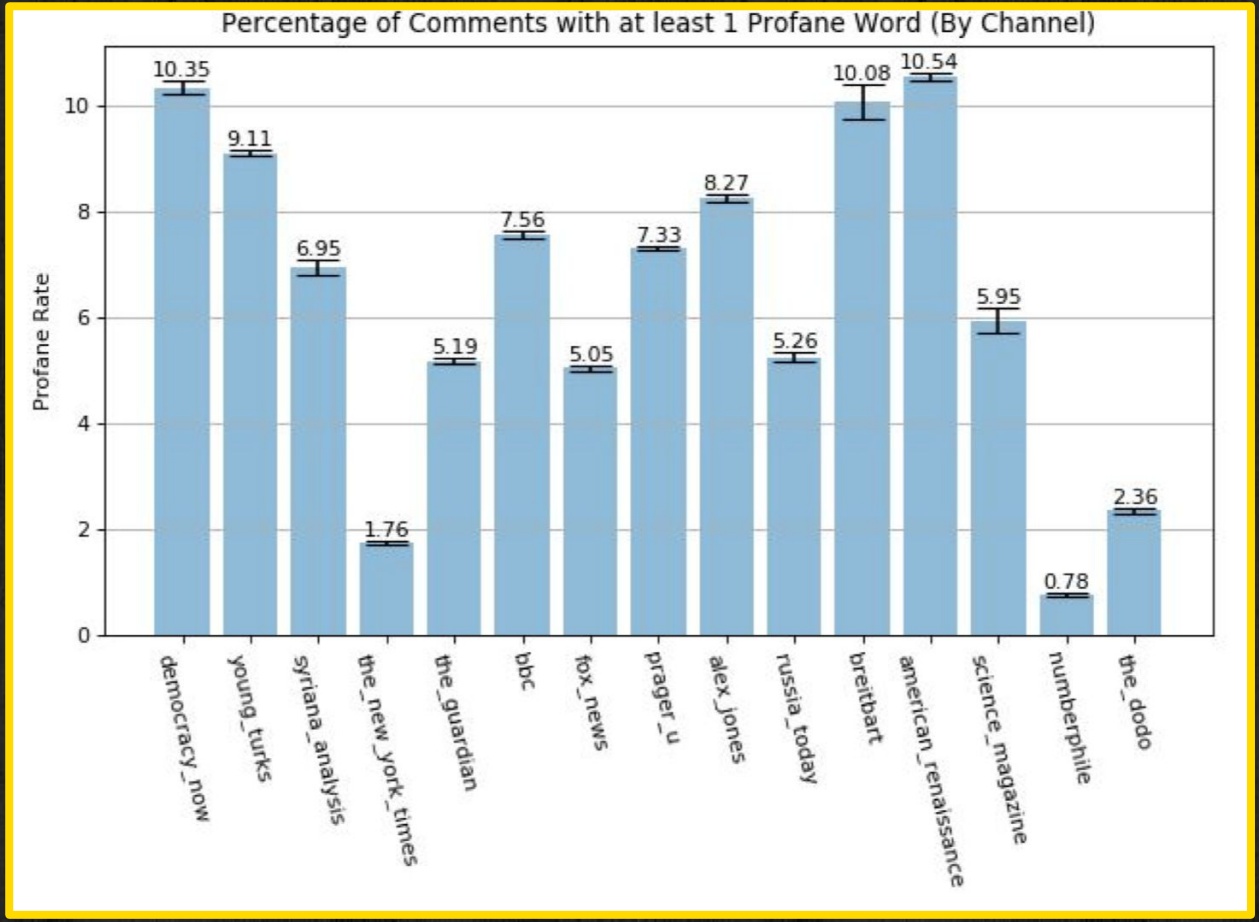
Results

Average Comment Length by Channel



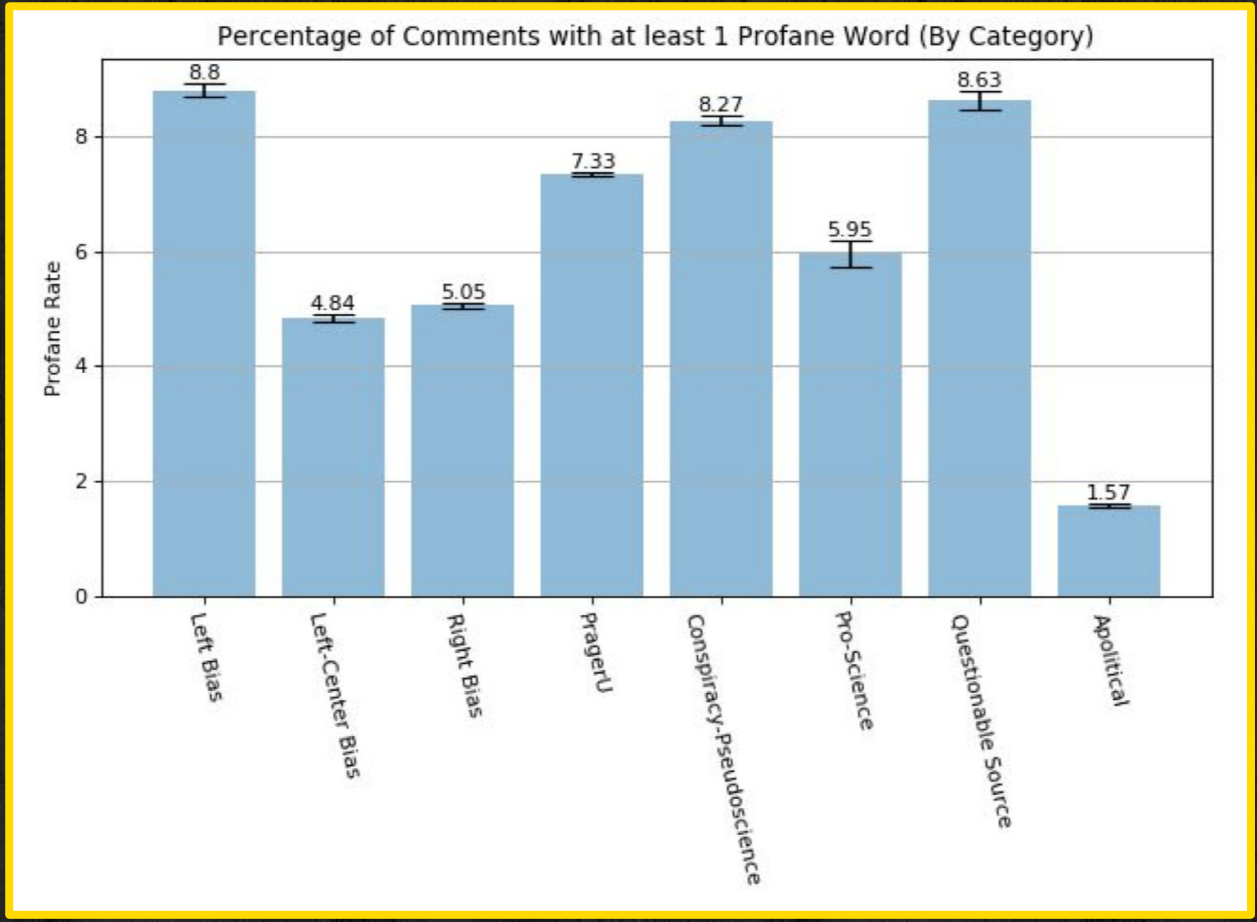
Results

Comments with Profanity by Channel



Results

Comments with Profanity by Category



Predicting the MBFC Category of a Video using Metadata

Machine Learning in Python Models:

- Decision Tree
- Random Forest
- Linear SVM
- SVM with RBF kernel

Predicting the MBFC Category of a Video using Metadata

Feature	Decision Tree Importance	Random Forest Importance
Num Comments	0.028195	0.067250
Average Direct Comment Like Count	0.071287	0.073852
Num Unique Authors	0.029831	0.076802
Average Thread Length	0.066847	0.079858
Average Comment Length	0.084224	0.085174
Dislike Count	0.124162	0.098351
Num Views	0.079629	0.107514
Profanity Rate	0.158240	0.121183
Comments Per View	0.133817	0.136023
Like Count	0.223767	0.153991

Notes

- Limited Data Set
- Mostly American Channels
- Different viewership across channels, Results may be different if a more comprehensive dataset would have been chosen

Future Research

- Why do certain comments receive greater traction?
- Why so some channels have longer ATL?
- How do discussions and arguments develop on YouTube?

Thanks!

Any questions?

Link to Paper

YouTube Chatter: Understanding Online Comments

Discourse on Misinformative and Political YouTube Videos