Test Summary

	BERTopic No Retweets	**BERTopic** With Retweets	**LSA** With Retweets
Model Coherence	58%	94%	42%
Tweet Volume	5,333 tweets	34,993 tweets	19,741 Took top 75% of tweets based on probability
Topics Selected based on tweet volume	BERTopic Labels: • subpoena_committee_subpoenaed_trump • **100% Coherent** • trumpcoupattempt_trump_day_trumpisgoingtojail • **75% Coherence** • and_forthepeople_peopleoverpolitics_standwithukraine • **67% Coherence** • rt_january6thcommitteehearings_watching_httpstcofcaejdpxxi • **67% Coherence**	BERTopic Labels: • both_they_backed_kevin • rudy_hatch_giuliani_meadows • demands_ja_deserves_unanimously • creating_hamill_correct_overthrowing • deploy_onthis_defense_sec	Human Created Labels: • Mitch_Kevin_Knew • Hamill_Democracy • Meadows_Guiliani_Stone • Subpoena_Trump
Tweet Volume By Topic	=> 140 tweets per topic	=> 1000 tweets per topic	
Takeaways	• Not high enough coherency for business confidence • See if there are specific topics that are scoring high to determine if we can uncover different topics when retweets removed.	• High coherency. High confidence that this model is accurately identifying different topics.	• Low Coherency. Don't use this model for business findings. • Lack of coherence reduces confidence in accuracy of human labeling

BERtopic vs KeyBERT

	BERTopic With Retweets	**KeyBERT** With Retweets
Human Coherence Score For Model	94%	97%
Tweet Volume	34,993 tweets	34,993 tweets
Topic Labels Top tweet volume	• both_they_backed_kevin • rudy_hatch_giuliani_meadows • demands_ja_deserves_unanimously • creating_hamill_correct_overthrowing • deploy_onthis_defense_sec	• mcconnell_mccarthy_senate_senators • democracy_january6thcomm_overthrowing_Hamill • trumpscoupplaybook_trump_giuliani_coup • subpoena_testify_oath_unanimously • miller_defense_deploy_sec
Takeaways	• Labels not intuitive	• Labels not intuitive

Annotator Observations

Don't exhaust annotators

- It's always good to use people that haven't been exposed to the content to get a clean read
- However, they can become bored and frustrated with the process when you give them several models to review.
- They may also become familiar with the data and that might bias their results. Be mindful of how many rounds of revised models you give to them to keep it fresh.