

[illegible]

1

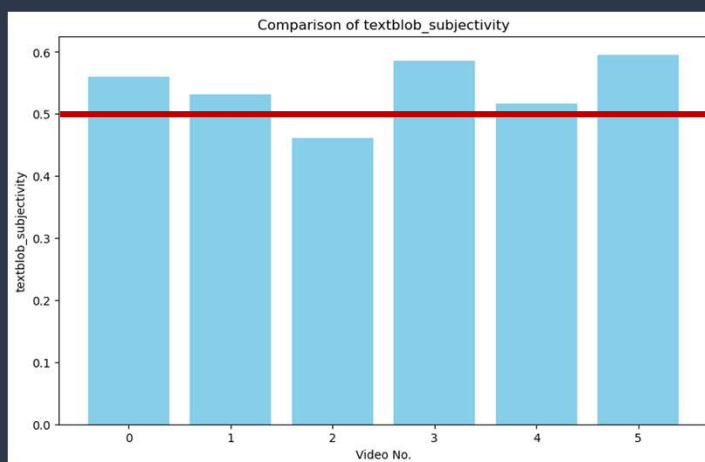
# WordCloud

[illegible]

3

# Subjectivity

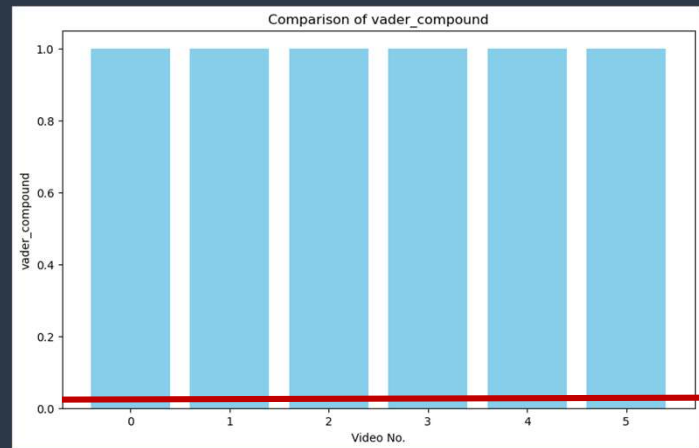
- Most videos are slightly subjective (above 0.5) but close to midway point between subjective and objective
- Some variation exists
  - For example, Video 2 is noticeably less subjective than the others



4

# Vader Scores

- Vader Compound Score combines all Vader scores: Positive, Neutral, and Negative into one score
- A score above 0.05 is considered to be positive
- For all videos, Wicked seems to garner overwhelmingly positive feedback



5

# Summary of Partition

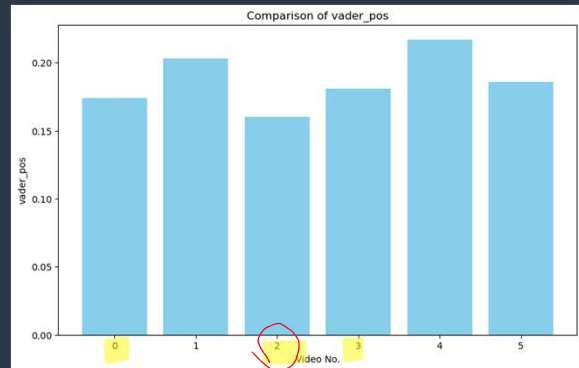
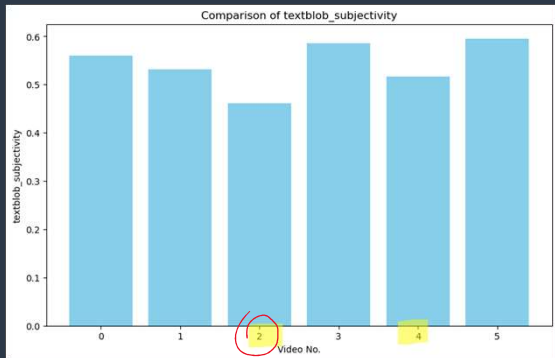
Method:

- By looking at the two charts, we first isolate each criteria separately and see which videos match it.
- Next, we find the intersecting matches between the two to find what meets our sorting criteria.
  - For instance, High Textblob\_Subjectivity and Low Vader Positive would be Negative-Subjective

Category	Video No.
Negative-Objective	2
Negative-Subjective	3
Positive-Objective	1,4
Positive-Subjective	0,5

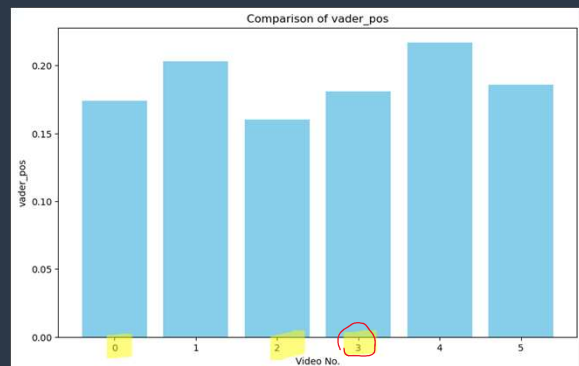
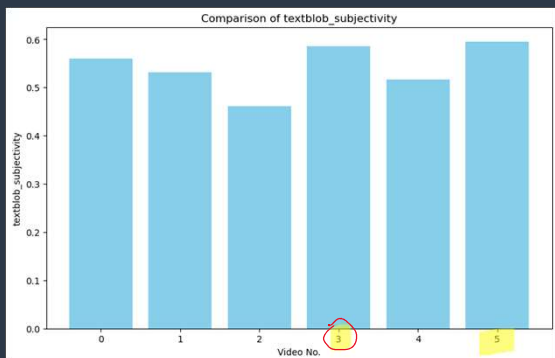
6

## Objective – Negative



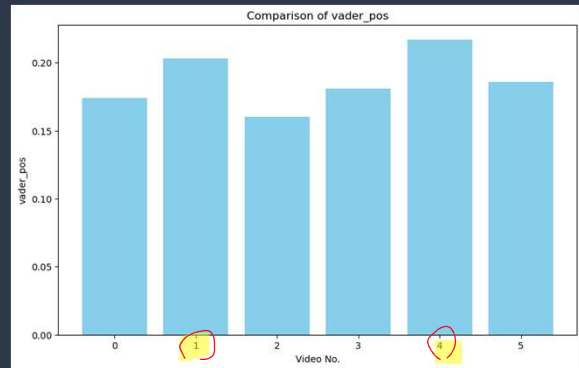
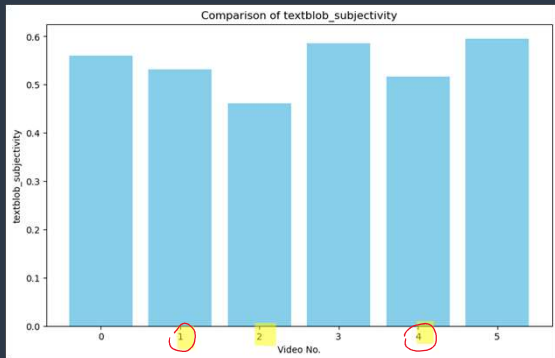
7

## Subjective - Negative



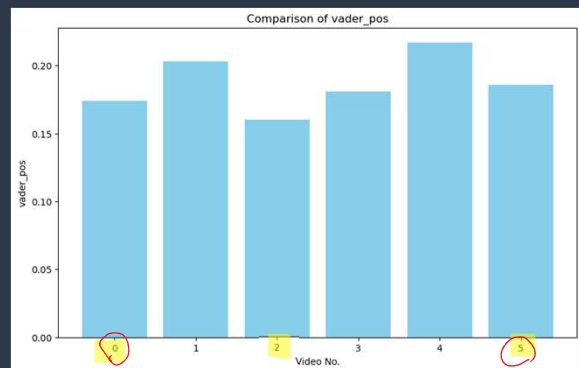
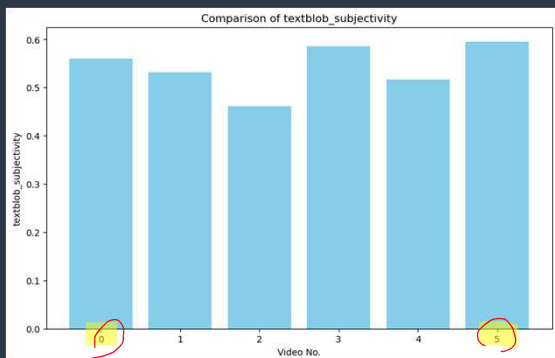
8

## Objective - Positive

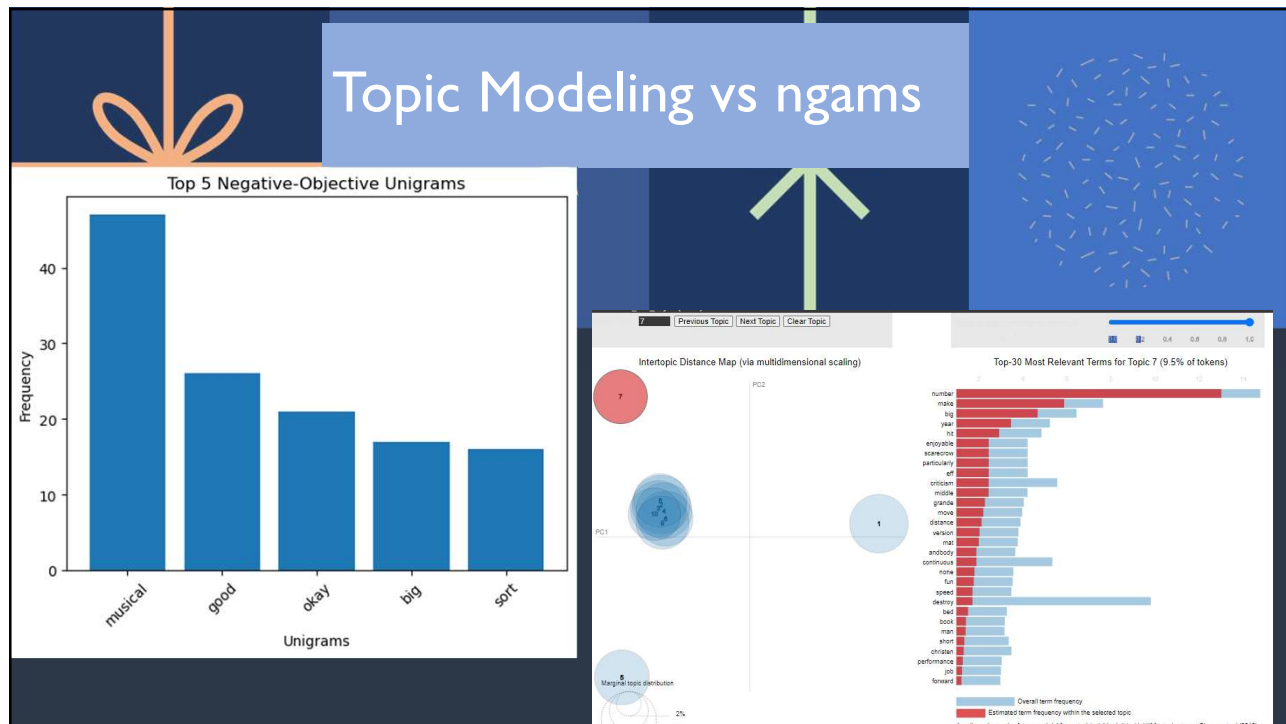


9

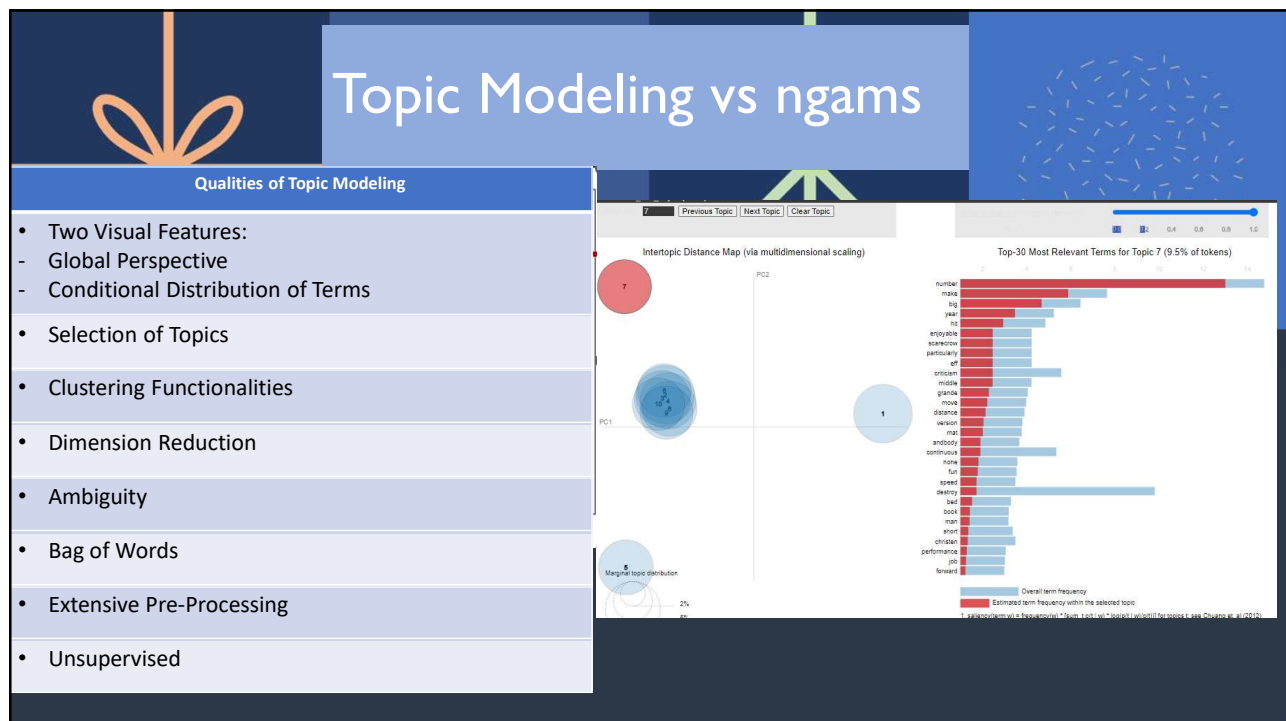
## Subjective - Positive



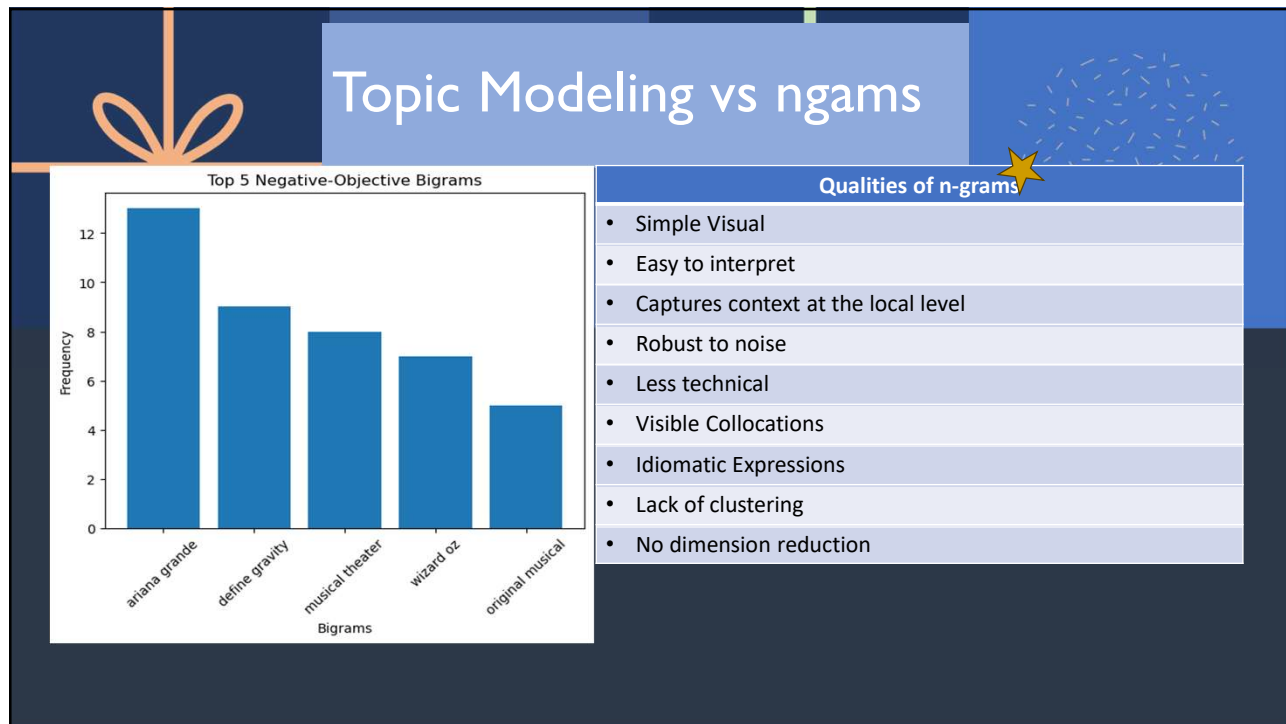
10



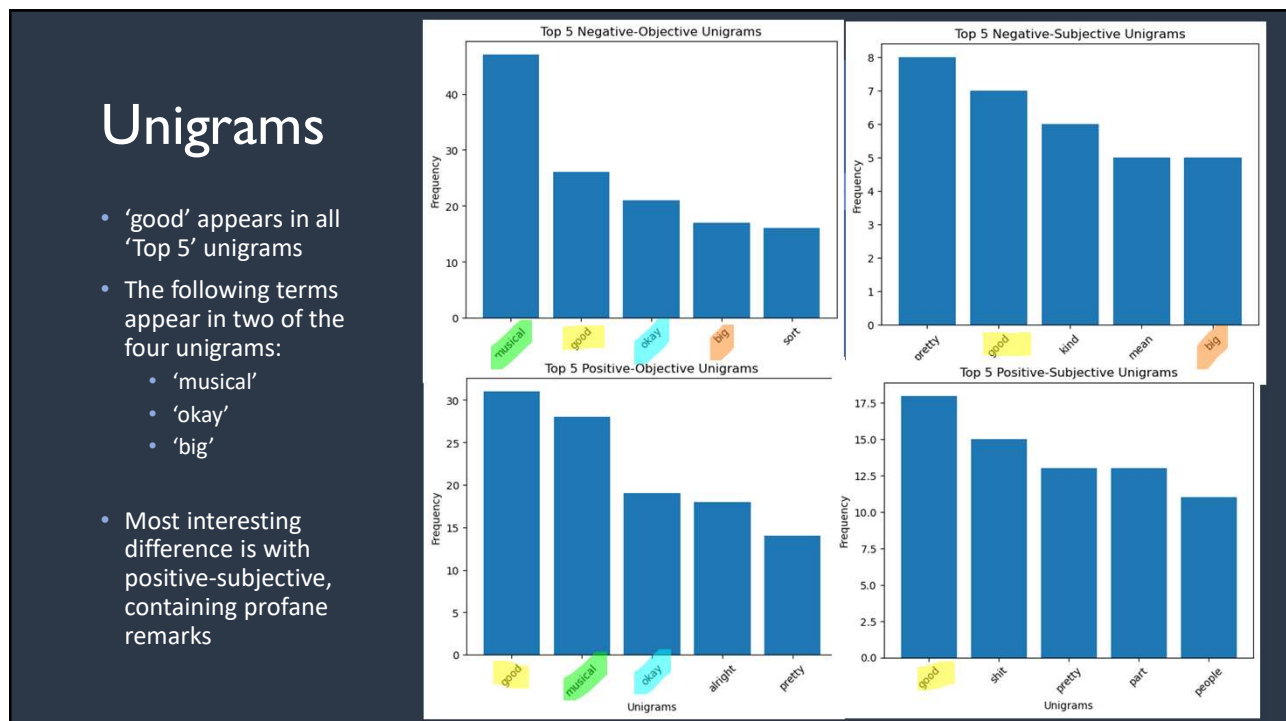
11



12

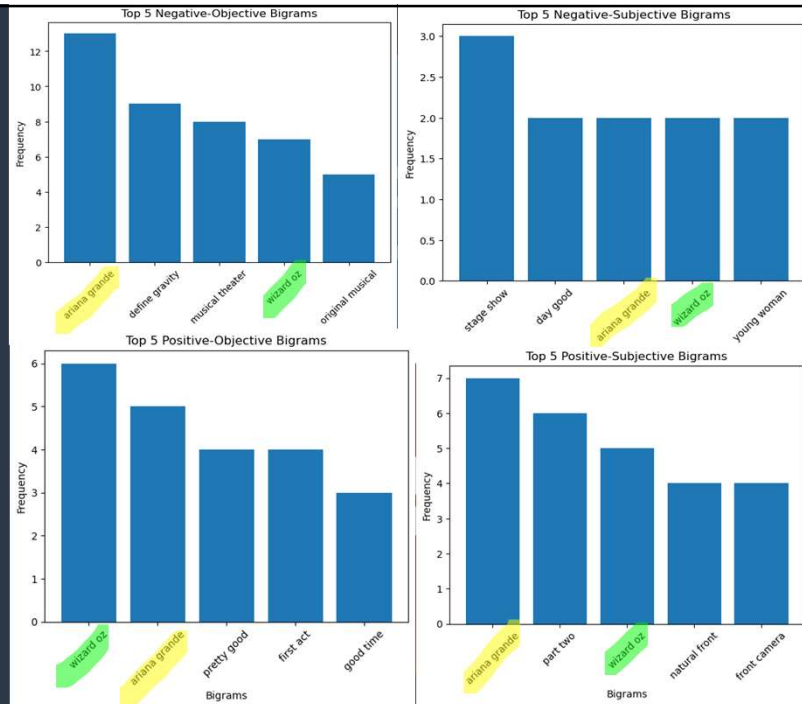


13



## Bigrams

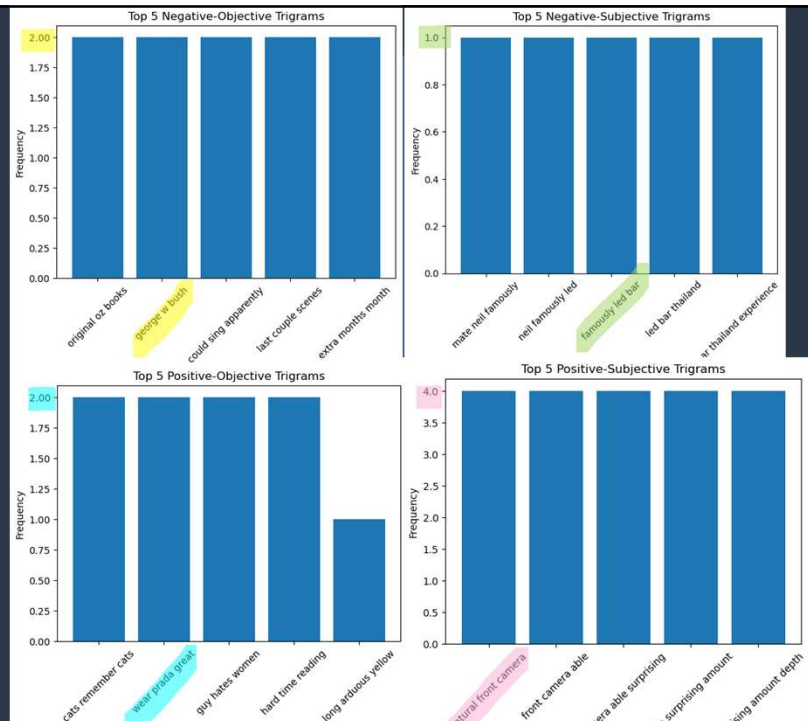
- 'Ariana grande' and 'wizard oz' are prevalent in all four bigrams
- Similar theater references for all except positive-subjective
  - 'musical theater'
  - 'first act'
  - 'stage show'
- Positive-Subjective focuses on the cinematic aspect
  - 'part two' instead of 'second act'
  - 'front camera'



15

## Trigrams

- Excluding Positive-Subjective, all other trigrams deviate off topic
  - Reflects podcasting hosts conversational style of video
  - Small sample size a potential factor as well
- Positive-Subjective stays focused on the film



16



# Topic Modeling

- An unsupervised classification of documents, similar to clustering, that finds natural groups of topics
- Is a generative probabilistic model
- Employs Bag-of-words
- Inherently stochastic
- Uses a similar method to PCA

Advantages	Disadvantages
Discovers hidden themes	Lack of context
Reduces dimensionality	Not <u>deterministic</u>
Organizes classifications to summarize the documents	No straightforward <u>method</u> of evaluating a model
Good for longer texts	Subjective interpretation

K-value: count of topics  
 'alpha': P(Document-Topic Density)  
 'eta': P(Word-Topic Density)

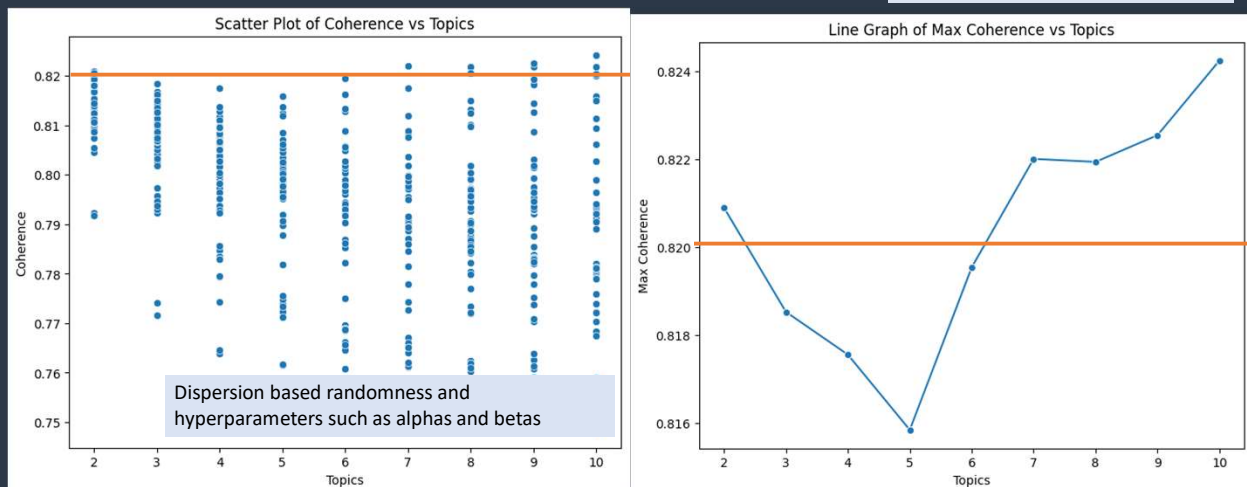
```
num_topics = 4

lda_model = gensim.models.LdaMulticore(corpus=corpus,
                                       id2word=id2word,
                                       num_topics=num_topics,
                                       random_state=100,
                                       chunksize=100,
                                       passes=10,
                                       alpha=0.91,
                                       eta=0.91)
```

17

# Optimized Coherence

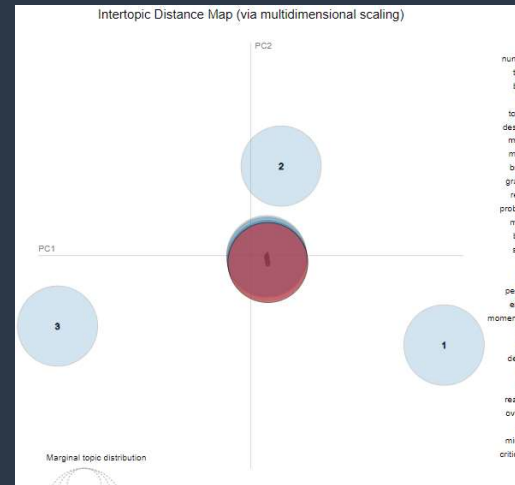
Coherence and Perplexity are inversely correlated. Coherence measures degree of topics are semantically similar. Perplexity measures how surprised model is with new data, given previous data. Optimizing perplexity is hard to interpret, so we optimize for coherence.



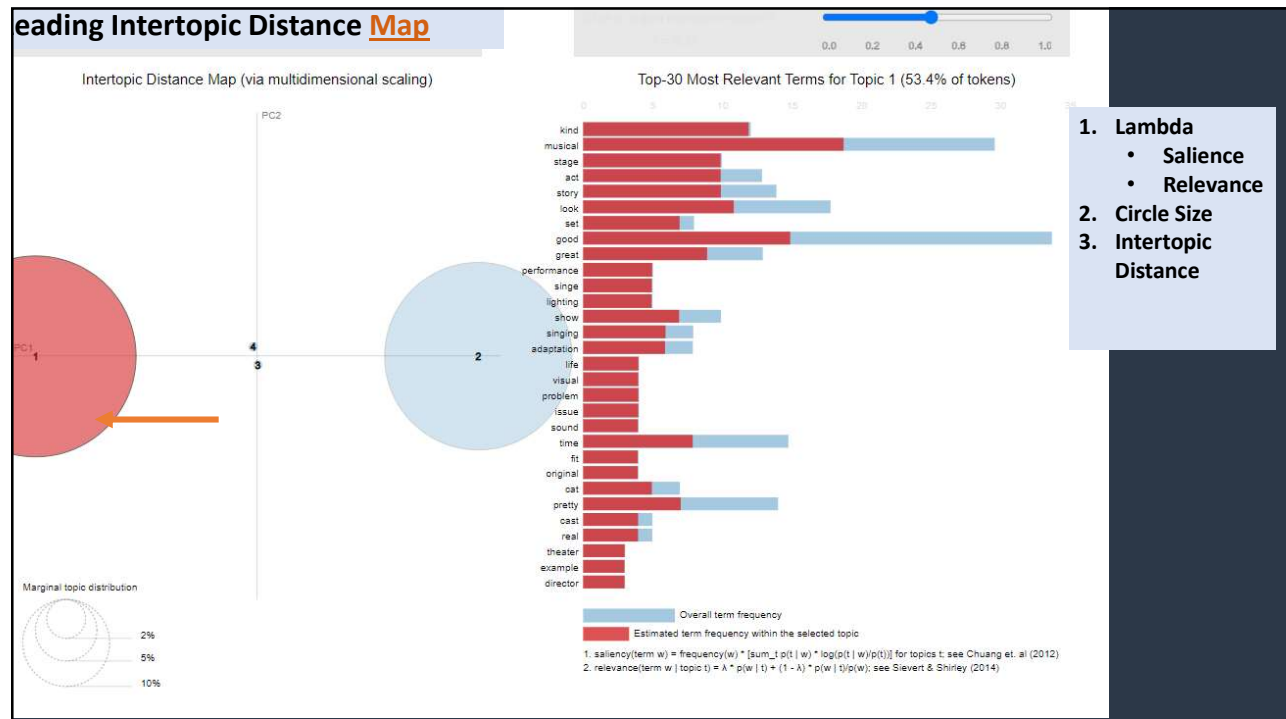
18

# Modeling Issues

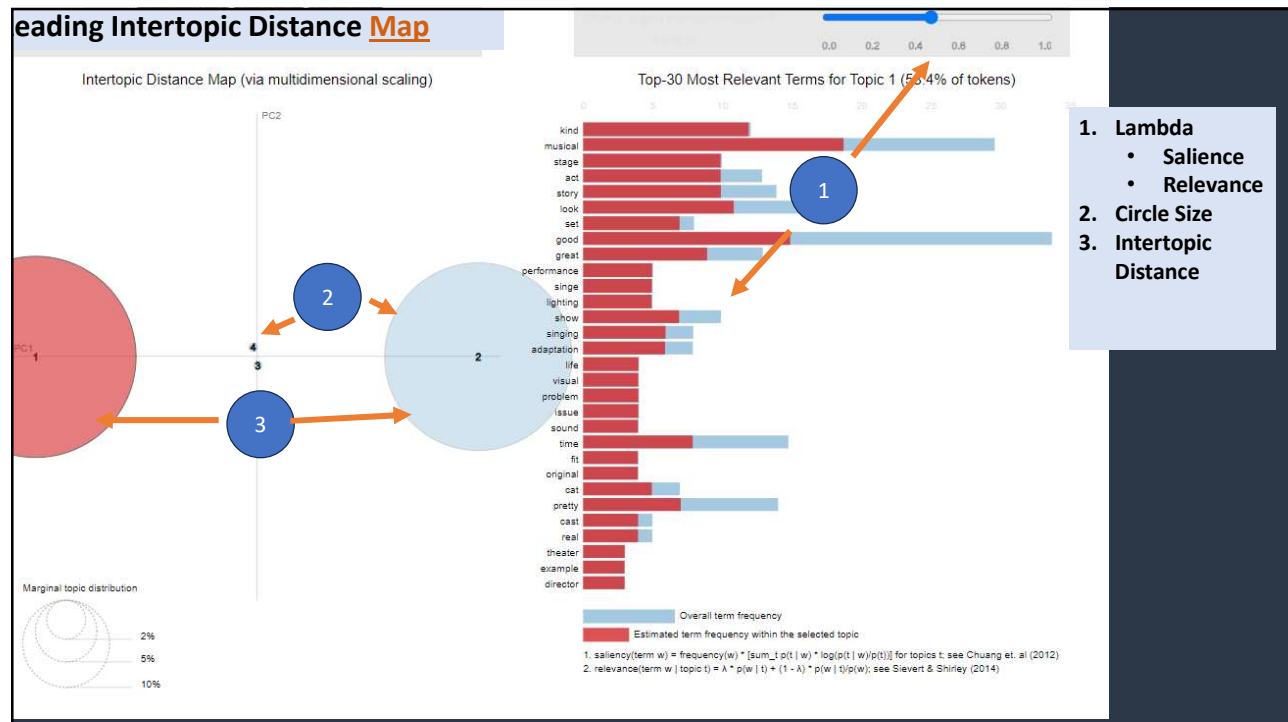
- Optimized coherence not always exact
  - Likely due to stochastic nature
- High  $k$  leads to duplicated topic
  - Can be visualized as overlapping circles
  - Lower  $k$ -values can yield similar coherence scores



19

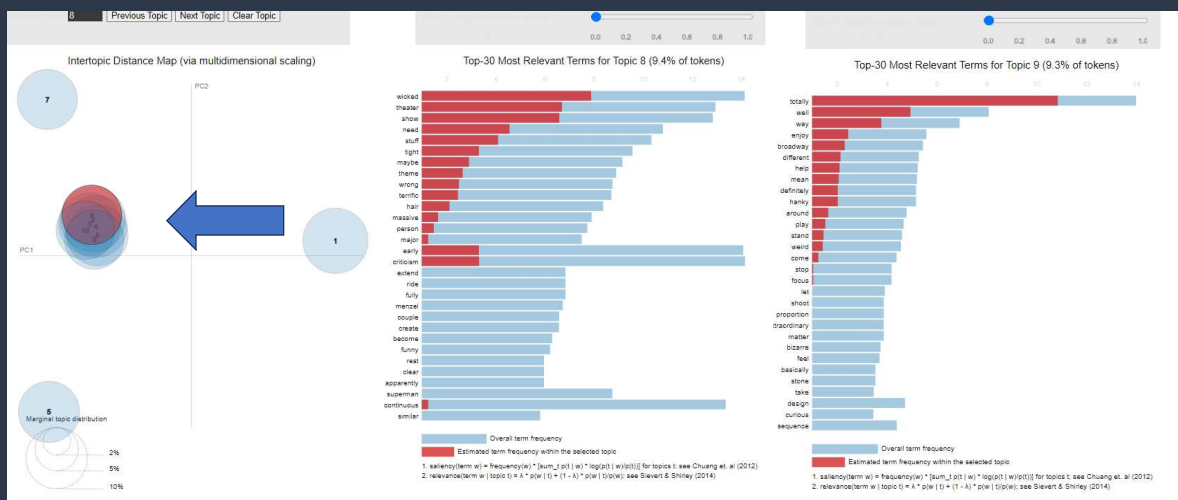


20



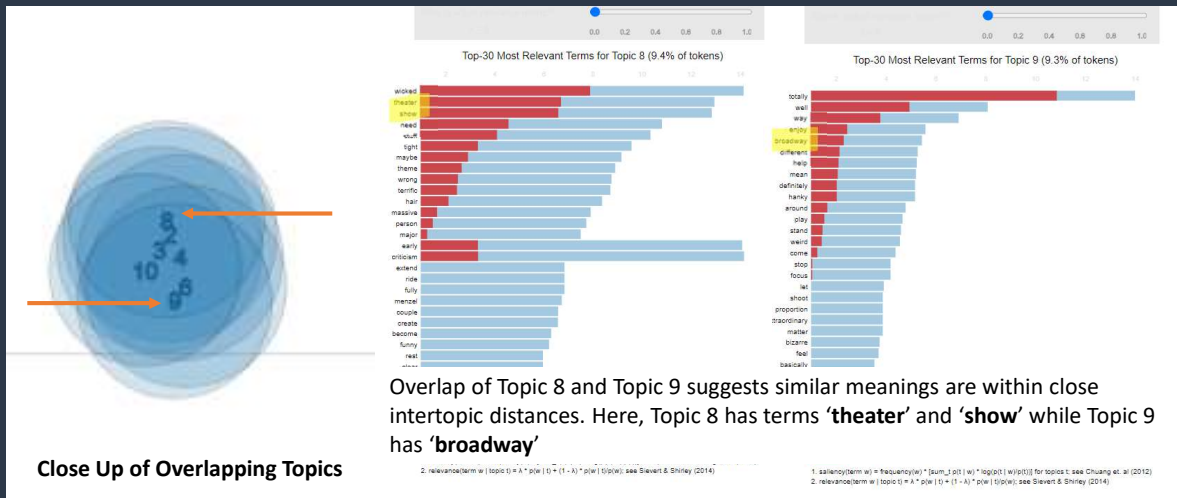
21

## LDA - pyLDAvis



22

# LDA - pyLDAvis



23

## LDA: Summary of Findings

	Negative-Objective	Negative-Subjective	Positive-Objective	Positive-Subjective
Optimum topics	2	4	4	3
Optimum word per topic	312	87	353	392
Perplexity	-7.04	-6.74	-8.48	-9.24
Coherence Score	0.81	0.82	0.27	0.23

Two terms you will want to understand when evaluating LDA models are:

- **Perplexity:** Lower the perplexity better the model.
- **Coherence:** Higher the topic coherence, the topic is more human interpretable.

24

## LDA: Summary of Findings

	Negative-Objective	Negative-Subjective	Positive-Objective	Positive-Subjective
Optimum topics	2	4	4	3
Optimum word per topic	312	87	353	392
Perplexity	-7.04	-6.74	-8.48	-9.24
Coherence Score	0.81	0.82	0.27	0.23

Expectation:  
Low Perplexity  
High Coherence

Highest Perplexity  
Highest Coherence

25

## LDA: Summary of Findings

	Negative-Objective	Negative-Subjective	Positive-Objective	Positive-Subjective
Optimum topics	2	4	4	3
Optimum word per topic	312	87	353	392
Perplexity	-7.04	-6.74	-8.48	-9.24
Coherence Score	0.81	0.82	0.27	0.23
Video Length	20 min (one video)	6 min (one video)	20 min (two 10-min video)	20 min (two 10-min video)

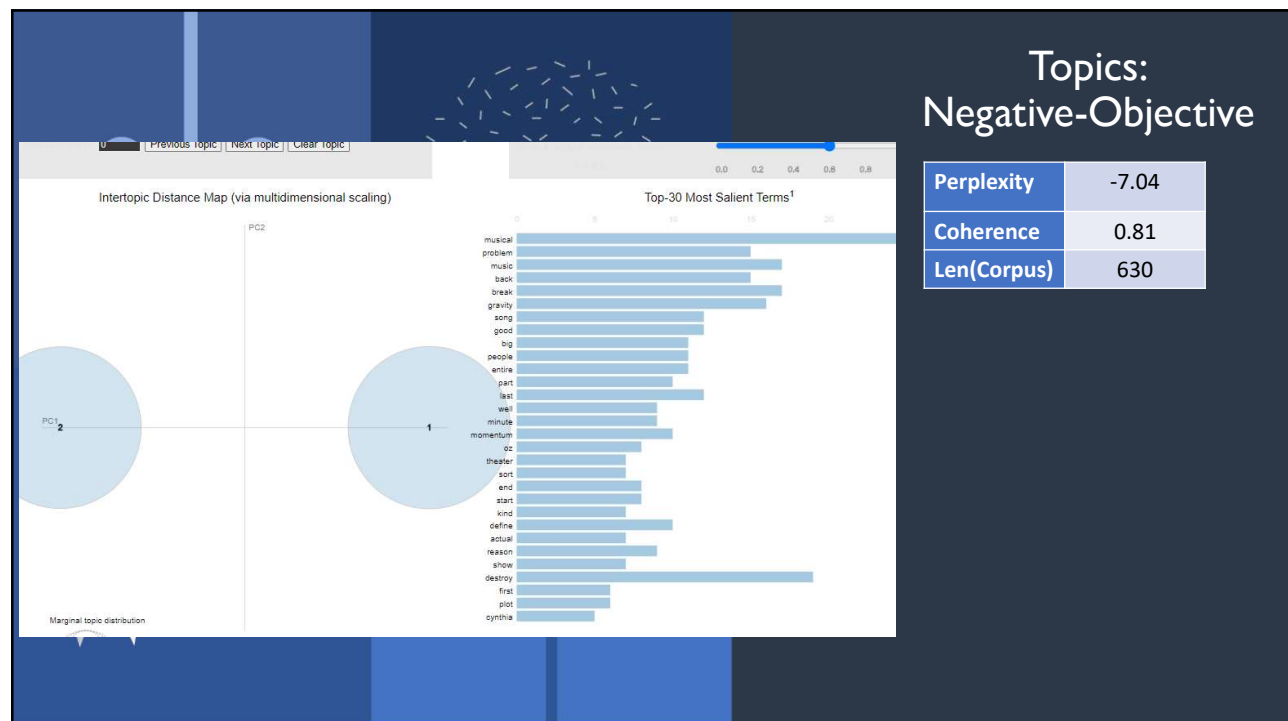
26

# LDA: Summary of Findings

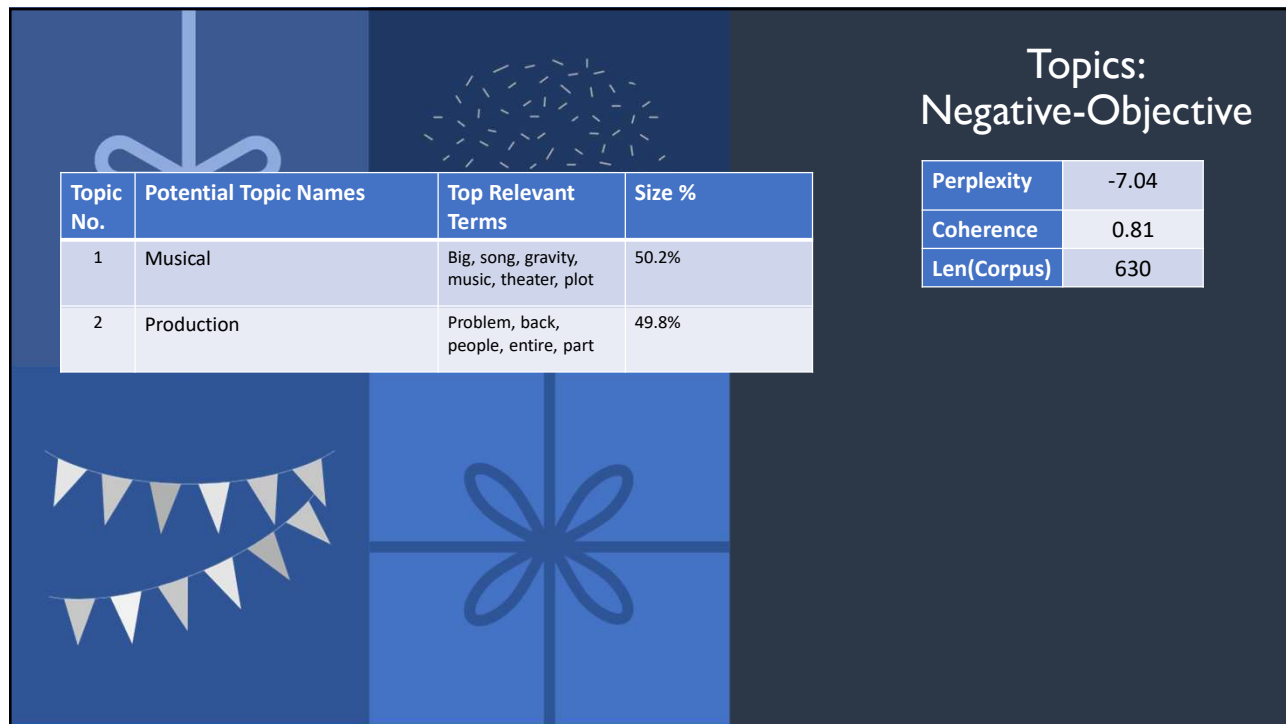
	Negative-Objective	Negative-Subjective	Positive-Objective	Positive-Subjective
Optimum topics	2	4	4	3
Optimum word per topic	312	87	353	392
Perplexity	-7.04	-6.74	-8.48	-9.24
Coherence Score	0.81	0.82	0.27	0.23
Video Length	20 min (one video)	6 min (one video)	20 min (two 10-min video)	20 min (two 10-min video)

- Topic size correlates with corpus size
- Coherence is sensitive to number of documents

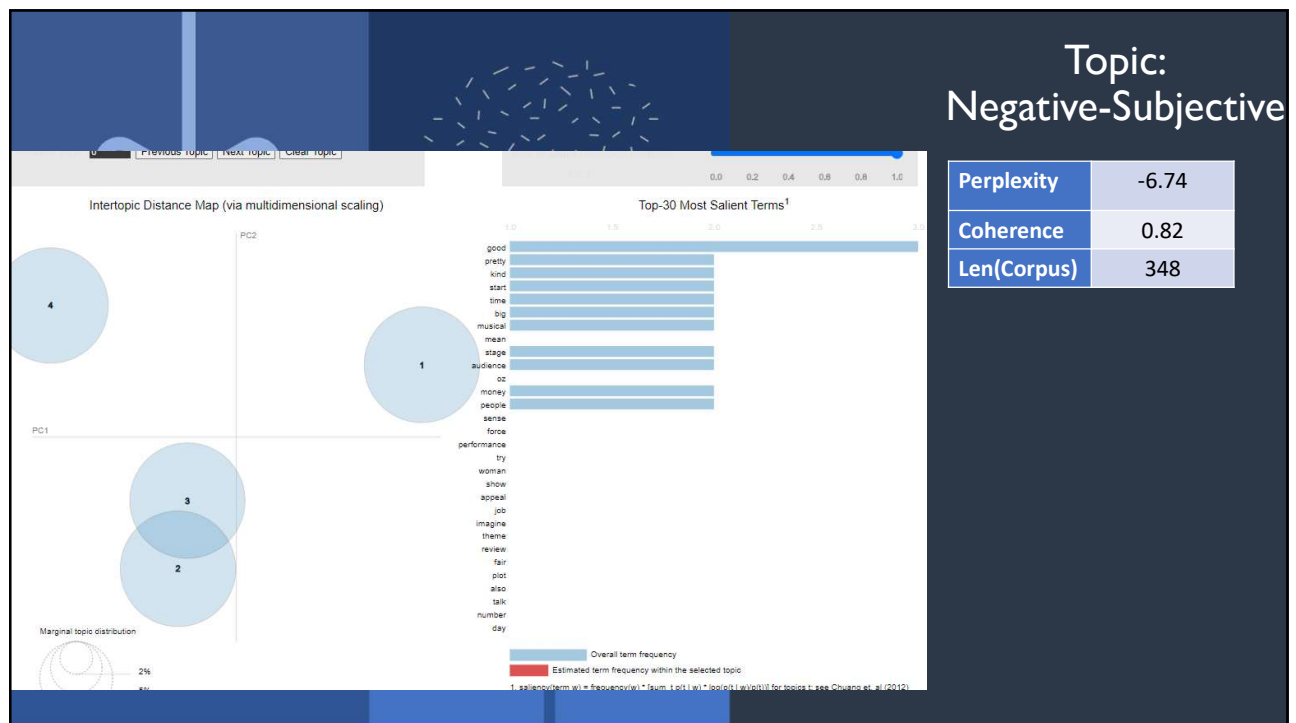
27



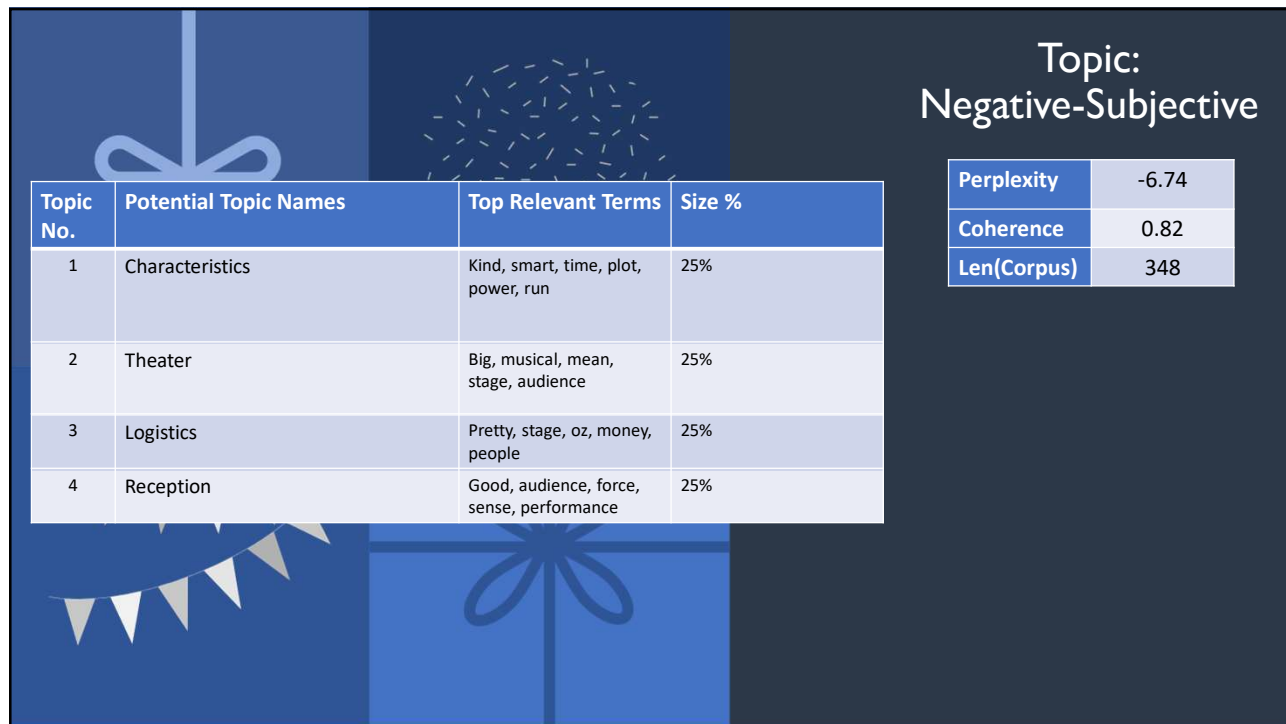
28



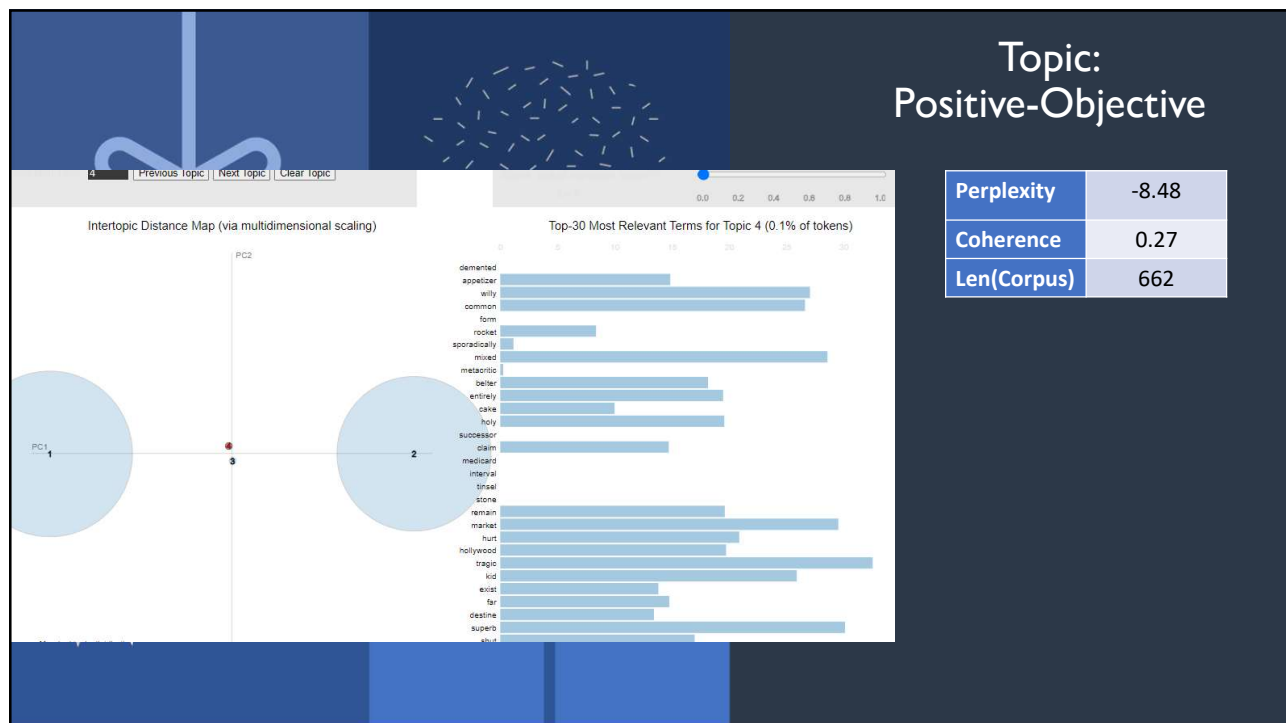
29



30

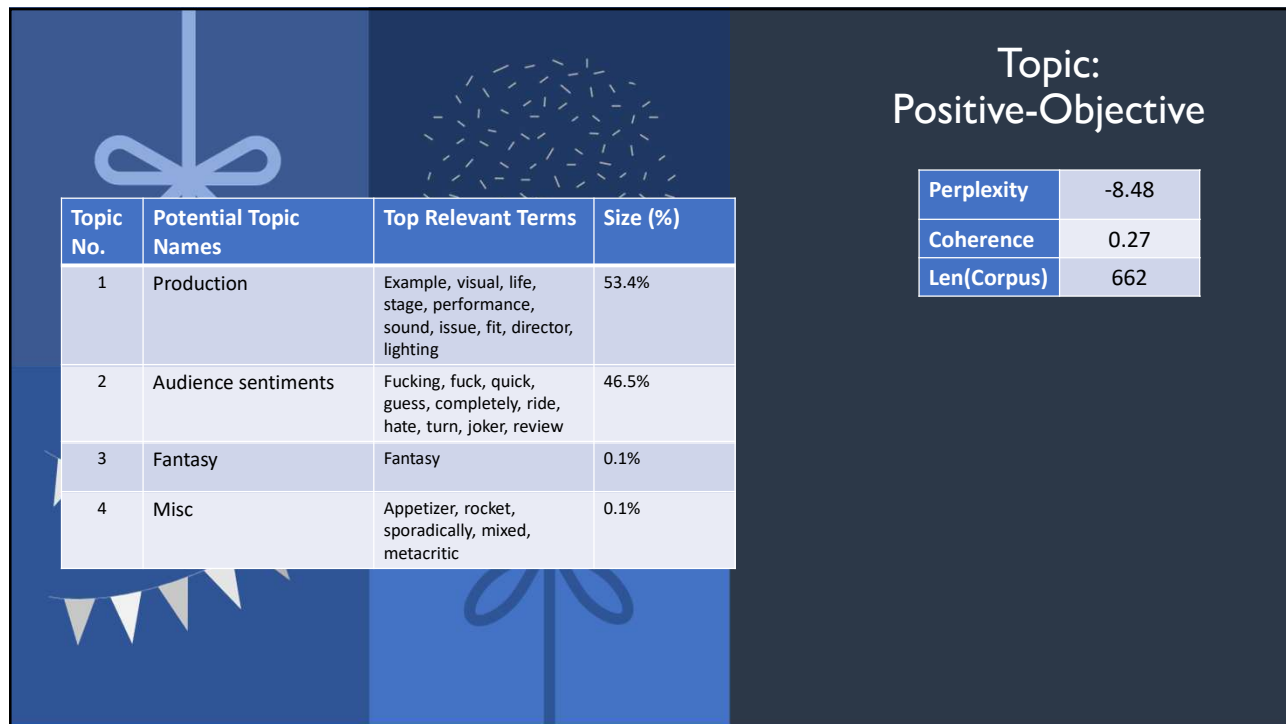


31

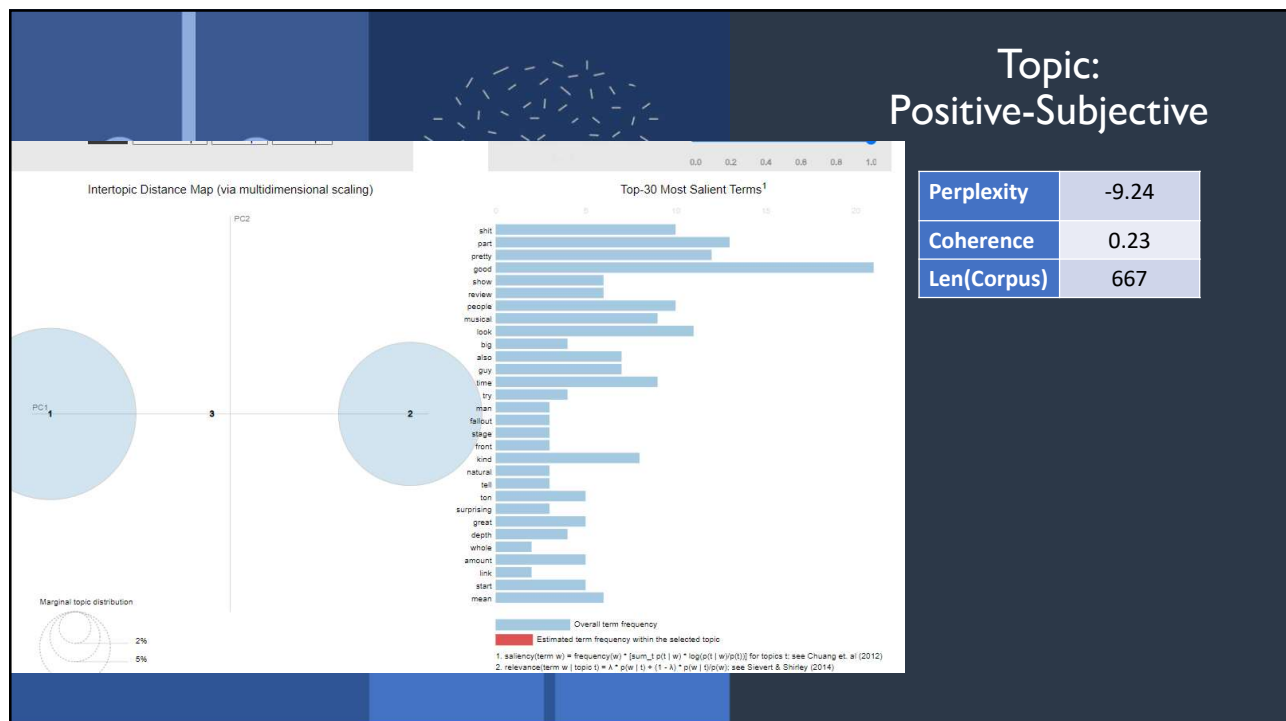


32





33



34

## Topic: Positive-Subjective

Topic No.	Potential Topic Names	Top Relevant Terms	Size (%)
1	Theater	show, stage, plot, social, whole	58.8%
2	Cinema	Enhance, tv, dolby, idea, cgi	41.2%
3	Acting	Face, amazing, motion, tilt	0.1%

Perplexity	-9.24
Coherence	0.23
Len(Corpus)	667

35

## Topics Side-by-Side Comparison

**Common to both is**

- Musical/Theater
- Production

Objective		Subjective	
Negative-Objective	Positive-Objective	Negative-Subjective	Positive-Subjective
Musical	Production	Characteristics	Theater
Production	Audience sentiments	Theater	Cinema
	Fantasy	Logistics	Acting
	Misc	Reception	

**Distinction is**

**Objective focuses on audience impressions, fantasy**

**Subjective focuses on acting, reception, character traits**

Objective reviews tend to focus on how the audience feels about the movie, while subjective reviews focus on acting and technicalities of the movies itself.

- Objective reviews evaluate based on how the audience feel about the film
  - Has a balance of good and bad remarks
- Subjective reviews only focuses on the film itself
  - Creates bias towards supporting the film

36

## Topics Side-by-Side Comparison

Positive		Negative	
Positive-Subjective	Positive-Objective	Negative-Subjective	Negative-Objective
Theater	Production	Characteristics	Musical
Cinema	Audience sentiments	Theater	Production
Acting	Fantasy	Logistics	
	Misc	Reception	

Similar to Objective-Subjective, Positive-Negative retains the same balance of distinctions. Positive is focused more on audience sentiments while Negative is focused on the film itself.

- May attributed to small sample size of the study
- Can also be explained by Negative review being overly technical in their review and lacking emotional appeal

37

## Additional resources

F, Sydney. "Getting to the Point with Topic Modeling: Part 3 - Interpreting the Visualization." *Maveryx Community*, 7 June 2023, [community.alteryx.com/t5/Data-Science/Getting-to-the-Point-with-Topic-Modeling-Part-3-Interpreting-the/ba-p/614992#:~:text=The%20intertopic%20distance%20map%20is,each%20topic%20across%20the%20dictionary.](https://community.alteryx.com/t5/Data-Science/Getting-to-the-Point-with-Topic-Modeling-Part-3-Interpreting-the/ba-p/614992#:~:text=The%20intertopic%20distance%20map%20is,each%20topic%20across%20the%20dictionary.)

P, Aditya. "How Do You Evaluate the Coherence and Perplexity of Your Topic Models with Gensim?" *Topic Modeling with Gensim: Coherence and Perplexity*, 25 Aug. 2023, [www.linkedin.com/advice/1/how-do-you-evaluate-coherence-perplexity](https://www.linkedin.com/advice/1/how-do-you-evaluate-coherence-perplexity).

Sievert, Carson, and Kenneth E. Shirley. *LDAvis: A Method for Visualizing and Interpreting Topics*, aclanthology.org/W14-3110.pdf. Accessed 10 Dec. 2024.

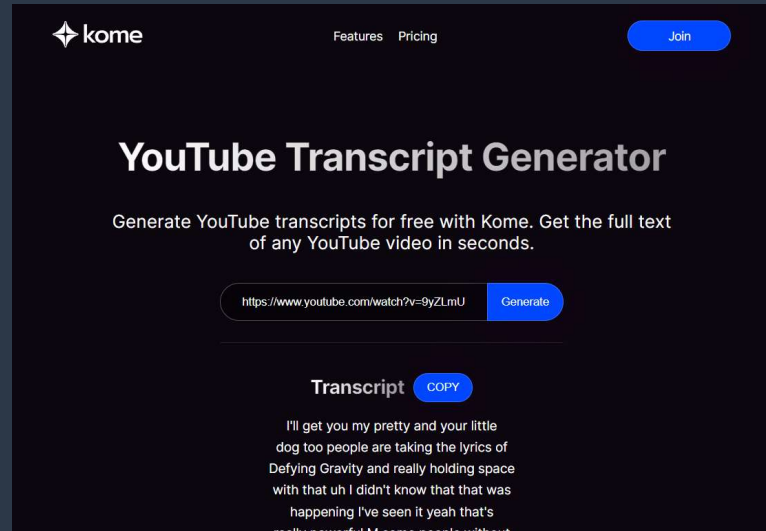
"Understanding LDA / Topic Modelling -- Too Much Topic Overlap." *Stack Overflow*, 1 Nov. 1962, [stackoverflow.com/questions/46326173/understanding-lda-topic-modelling-too-much-topic-overlap](https://stackoverflow.com/questions/46326173/understanding-lda-topic-modelling-too-much-topic-overlap).

AIML.com. "What Is Topic Modeling? Discuss Key Algorithms, Working, Applications, and the Pros and Cons." *AIML.Com*, 17 Oct. 2024, [aiml.com/what-is-topic-modeling/#:~:text=Lack%20of%20Context%3A%20Topic%20modeling,ability%20to%20capture%20nuanced%20meanings.&text=Ambiguity%3A%20Some%20words%20may%20appear,interpreting%20topics%20and%20their%20boundaries.](https://aiml.com/what-is-topic-modeling/#:~:text=Lack%20of%20Context%3A%20Topic%20modeling,ability%20to%20capture%20nuanced%20meanings.&text=Ambiguity%3A%20Some%20words%20may%20appear,interpreting%20topics%20and%20their%20boundaries.)

38

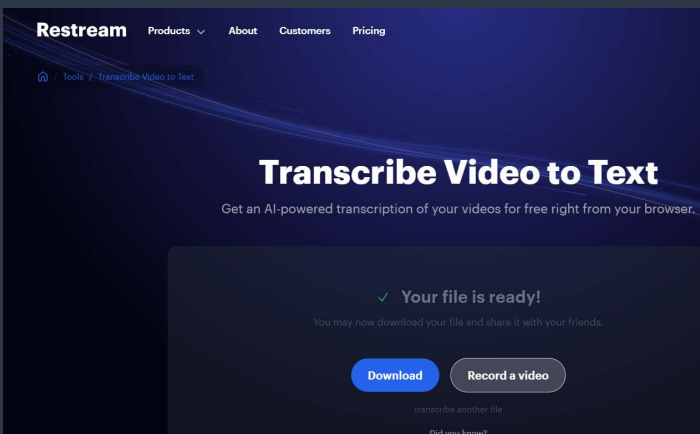
# Kome.ai

- “All-in-one tool for knowledge”
- Caters to individuals
- Primarily a transcription service
  - Available as “copy”
- Marketing convenience
  - Bookmark feature
  - Summary capabilities
  - “email” extraction feature
- Wants to be your companion
  - Browser Extension
- Free to use
- Sign up for account customization
- Basic plan is at \$5/mo



39

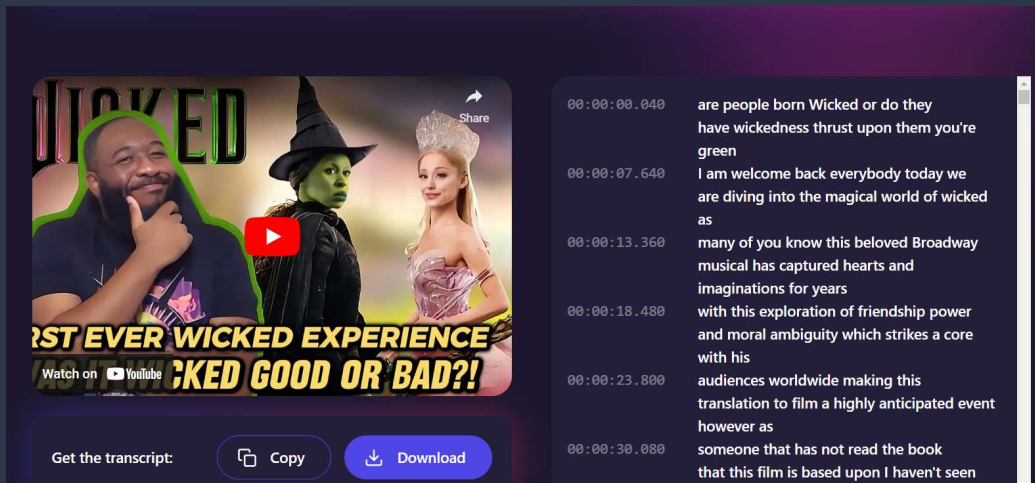
# Restream.io



- “We help stories be seen and heard”
- Caters to Video Streamers
- Considerably larger company
- Provides multi-streaming capabilities, video editing studio, and technical support service
- Transcription was a free feature
  - Download as txt
- Account sign up required for additional functionalities
- Team environment available
- Enterprise pricing available
- Starting price at \$16/mo

40

# Tactiq.io



The screenshot shows the Tactiq.io interface. On the left, there is a video player with a thumbnail featuring a man and two women from the musical 'Wicked'. The video title is 'FIRST EVER WICKED EXPERIENCE' and the subtitle is 'WICKED GOOD OR BAD?!'. Below the video player, there are buttons for 'Get the transcript:', 'Copy', and 'Download'. On the right, there is a transcript of the video content.

00:00:00.040 are people born Wicked or do they have wickedness thrust upon them you're green

00:00:07.640 I am welcome back everybody today we are diving into the magical world of wicked as

00:00:13.360 many of you know this beloved Broadway musical has captured hearts and imaginations for years

00:00:18.480 with this exploration of friendship power and moral ambiguity which strikes a core with his

00:00:23.800 audiences worldwide making this translation to film a highly anticipated event however as

00:00:30.080 someone that has not read the book that this film is based upon I haven't seen

41

# Tactiq.io

- "Focus on the Meeting, let AI handle the notes."
- Caters to businesses
- Primarily a meetings productivity app
- Selling point is transcription of live meetings
- Integrates with Google Meet, Zoom, Microsoft Teams
- Browser Extension available
- Video transcription is free
  - Export as txt
  - Timed transcription
- Team environment available
- Basic subscription starts at \$12/mo
- Enterprise options available

00:00:00.040 are people born Wicked or do they have wickedness thrust upon them you're green

00:00:07.640 I am welcome back everybody today we are diving into the magical world of wicked as

00:00:13.360 many of you know this beloved Broadway musical has captured hearts and imaginations for years

00:00:18.480 with this exploration of friendship power and moral ambiguity which strikes a core with his

00:00:23.800 audiences worldwide making this translation to film a highly anticipated event however as

00:00:30.080 someone that has not read the book that this film is based upon I haven't seen

42

# Transcription Services

		<b>Restream</b>	
Youtube-Compatible	✓	X	✓
Browser Extension	✓	X	✓
Use of AI	✓	✓	✓
Video Editing	X	✓	X
Bookmark Capability	✓	X	X
Timed Transcription	X	X	✓
Export Transcription	X	✓	✓
Sign Up Required	X	X	X
Account Features	✓	✓	✓
Team Features	X	✓	✓
Basic Plan	\$5/mo	\$16/mo	\$12/mo

43

# Transcription Services

		<b>Restream</b> 	
Youtube-Compatible	✓	X	✓
Browser Extension	✓	X	✓
<b>Use of AI</b>	✓	✓	✓
Video Editing	X	✓	X
Bookmark Capability	✓	X	X
Timed Transcription	X	X	✓
Export Transcription	X	✓	✓
Sign Up Required	X	X	X
<b>Account Features</b>	✓	✓	✓
Team Features	X	✓	✓
Basic Plan	\$5/mo	\$16/mo	\$12/mo

44