

The word "JEOPARDY!" is rendered in its iconic, bold, silver 3D block letters. The letters have a metallic sheen and are set against a vibrant background that transitions from a deep blue at the top to a bright orange and red at the bottom. The background is decorated with numerous white starburst and star patterns of varying sizes, creating a cosmic or night-sky effect.

**JEOPARDY!**

The words "TRIVIA TOPIC MODELING" are written in a clean, white, sans-serif font. The text is centered horizontally and has a subtle drop shadow, making it stand out against the colorful, patterned background. Below the text, there is a faint, white grid of intersecting lines that adds a technical or digital feel to the overall design.

**TRIVIA TOPIC MODELING**

BEFORE & AFTER	LITERATURE	SCIENCE	AMERICAN HISTORY	POTPOURRI	SAY "WH"
\$200	\$200	\$200	\$200	\$200	\$200
\$400	\$400	\$400	\$400	\$400	\$400
\$600	\$600	\$600	\$600	\$600	\$600
\$800	\$800	\$800	\$800	\$800	\$800
\$1000	\$1000	\$1000	\$1000	\$1000	\$1000

BEFORE & AFTER · \$200

The rash-causing 3-leafed shrub is  
part of a collegiate group with  
Dartmouth & Harvard.

[← BACK TO PANEL](#)

BEFORE & AFTER · \$1000

Botticelli painting in which Serena's  
sister is brought forth on a shell.

[← BACK TO PANEL](#)

# GOAL

- Categorize corpus of Jeopardy! questions into meta-categories through topic modeling.
- Deploy a responsive study app to help people interested in competing on Jeopardy!



# DATA

The original dataset contains ~220K Jeopardy questions from 1984 to 2012. Each row contains:

Air Date	Round	Category	Value	Question	Answer
----------	-------	----------	-------	----------	--------

# TOOLS USED

Pandas, NLTK, Sklearn, StreamLit, WordCloud

## PROCESS

1

Read in  
Text Data

2

Preprocess  
Text

3

Featurize  
Text using  
TF-IDF

4

Unsupervised  
NMF Model to  
Uncover  
Hidden Topics

5

Results in  
meta-  
Categories

# PROCESS

1

Read in  
Text Data

2

Preprocess  
Text

3

Featurize  
Text using  
TF-IDF

4

Unsupervised  
NMF Model to  
Uncover  
Hidden Topics

5

Results in  
meta-  
Categories

## ORIGINAL

{  
"HISTORY"  
4 treaties to  
mitigate the  
horrors of war  
were ...  
Geneva"}  
}



Down-sample  
Lemmatization  
POS Nouns  
Stopword removal:  
['seen here', '<html>', 'clue crew']

## PROCESSED

["history"  
"treaty" "horror"  
"war" ...  
"geneva"]



# PROCESS

1

Read in  
Text Data

2

Preprocess  
Text

3

Featurize  
Text using  
TF-IDF

4

Unsupervised  
NMF Model to  
Uncover  
Hidden Topics

5

Results in  
meta-  
Categories

## Document-Term Matrix

Jeopardy! Clues (Docs)	Terms			
	act	actor	battle...	world
0	0.0	0.0	0.31	0.14
1	0.14	0.09	0.0	0.0
...				
101715	0.0	0.0	0.0	0.24

=

## H Matrix

Hidden Topics

Terms	0.0	...	0.7
	0.14	...	0.0
	...	...	...
	0.0	...	0.8

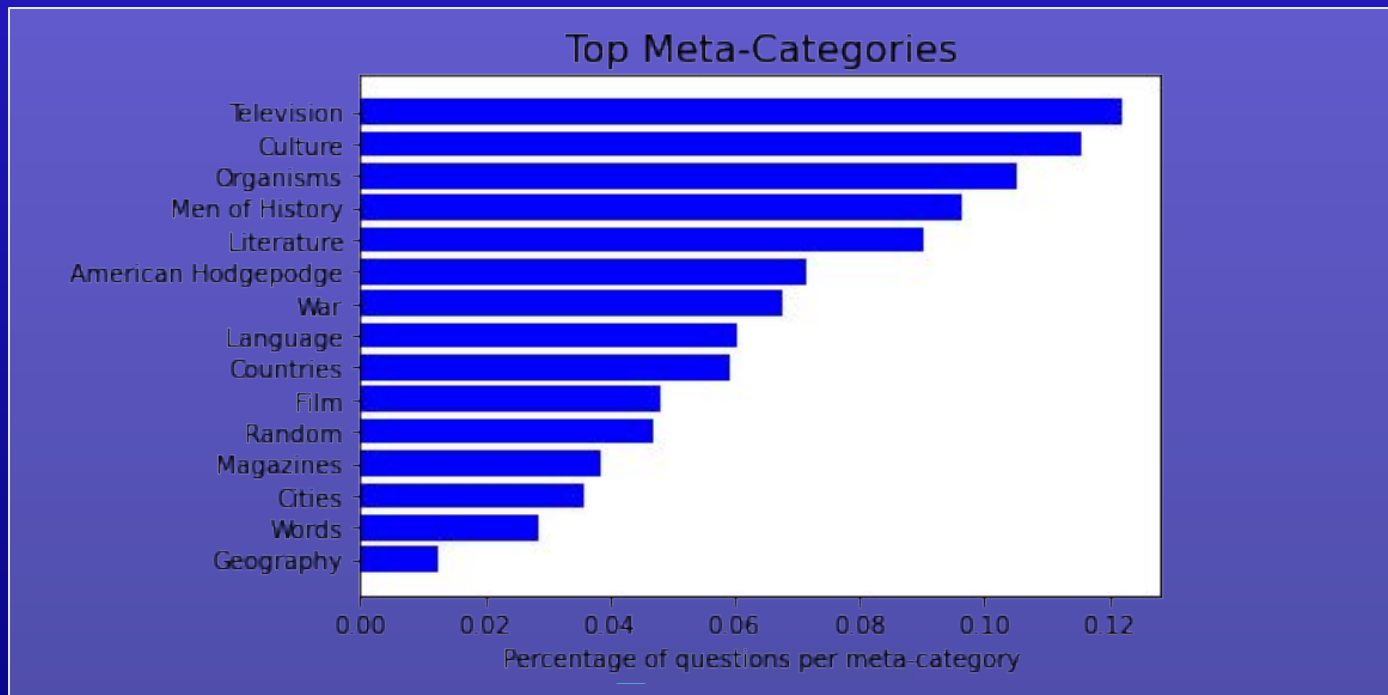
×

## W Matrix

Jeopardy! Clues

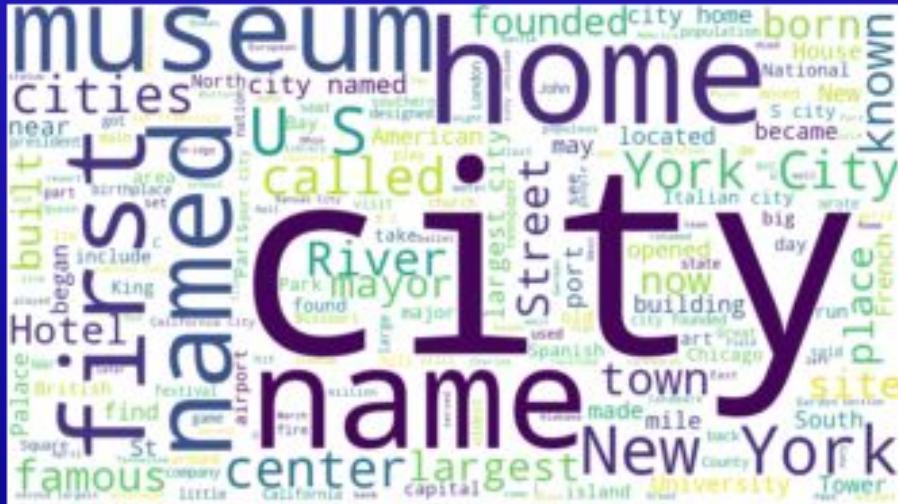
Hidden Topics	0.0	0.0	...	0.14
	...	...	...	...
	0.0	0.0	...	0.24

# EXPLORING THE DATA



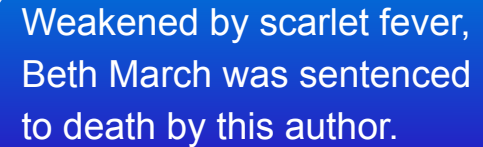


# CITIES



4 treaties to mitigate the horrors of war were signed in this city in August, 1949.

The last of the 13 colonies to be founded, its 'Mother City', Savannah, was settled in 1733.



In a 14-line poem  
Wordsworth wrote "Scorn  
not" this form of poetry.

# FILM



1940's "The Shop Around the Corner" was the basis for this 1998 Meg Ryan-Tom Hanks romantic comedy

## ‘The Apostle’

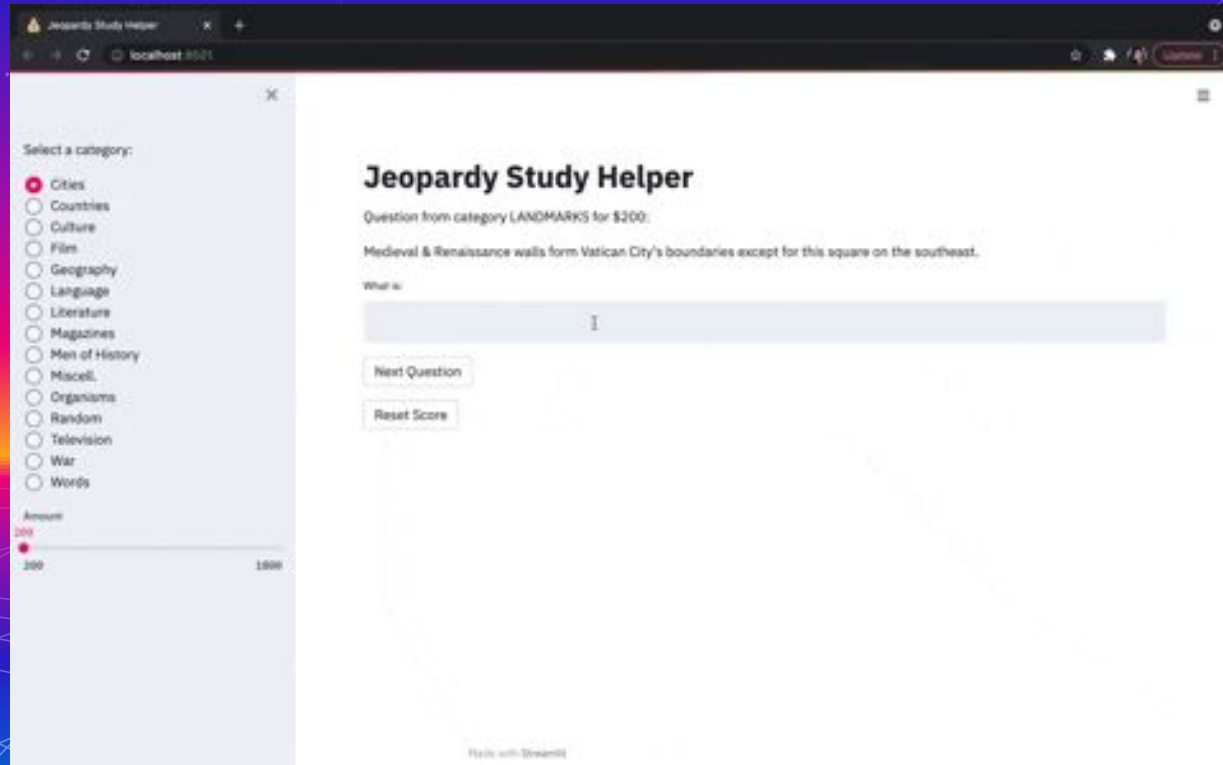
# TELEVISION



Twin brothers named Dromio are servants to twin brothers name Antipholus in this highly confusing play.

## Something to Talk About (1991)

# STREAMLIT APP DEMO



# CONCLUSIONS

- Jeopardy's corpus lends itself nicely to NLP techniques.
- Questions highly dependent on current times, can get outdated quickly.
- N=15 topics produced nicely delineated topics, but additional processing required
- Recurrent themes suggest a recommendation algorithm can help students focus their studies



# NEXT STEPS

- Handle images, video to expand question set
- Restrict questions to more recent years, add questions after 2012
- Implement CorEx-based topic modeling with anchor words
- Recommendation algorithm to recommend subjects for study based on prior performance
- Fix bugs in web app, deploy with Heroku