Text Analysis Report:
marvel_cinematic_universe_wikipedia

1. Basic Statistics

Metric	Value
Word Count	460
Sentence Count	29
Character Count	2912
Unique Word Count	221
Avg Word Length (chars)	5.11
Avg Sentence Length (words)	15.86
Avg Sentence Length (chars)	100.41
Sentence Length Std Dev (words)	13.84
Lexical Diversity (TTR)	0.4804
MATTR (Window=50)	0.7806

## 2. Entropy & Complexity
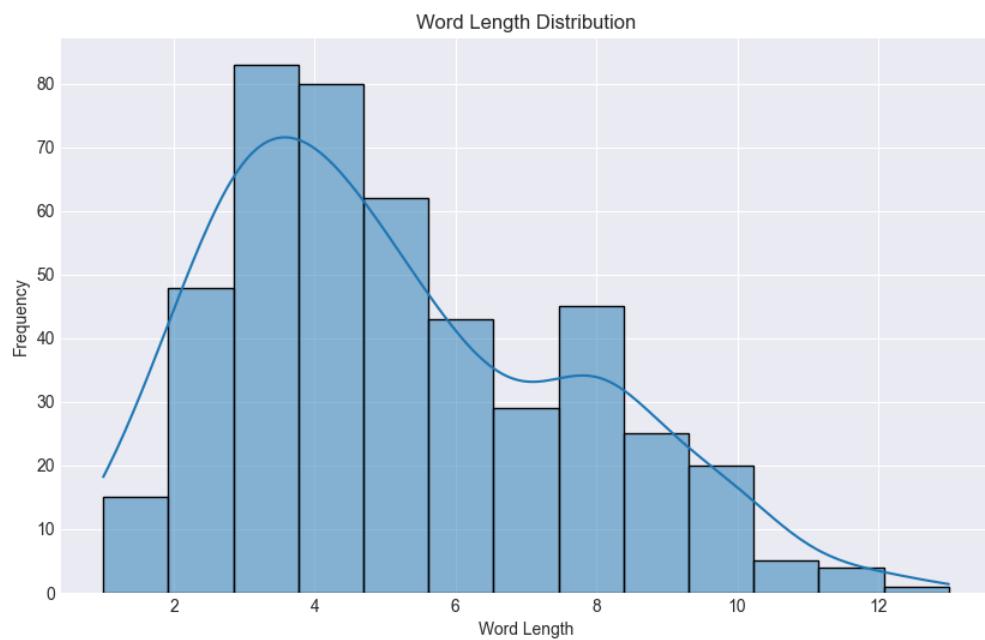
Metric	Value
Shannon Entropy (Words)	7.1731 bits
Shannon Entropy (Chars)	4.7143 bits
Function Word Ratio	0.3217

3. Readability Scores
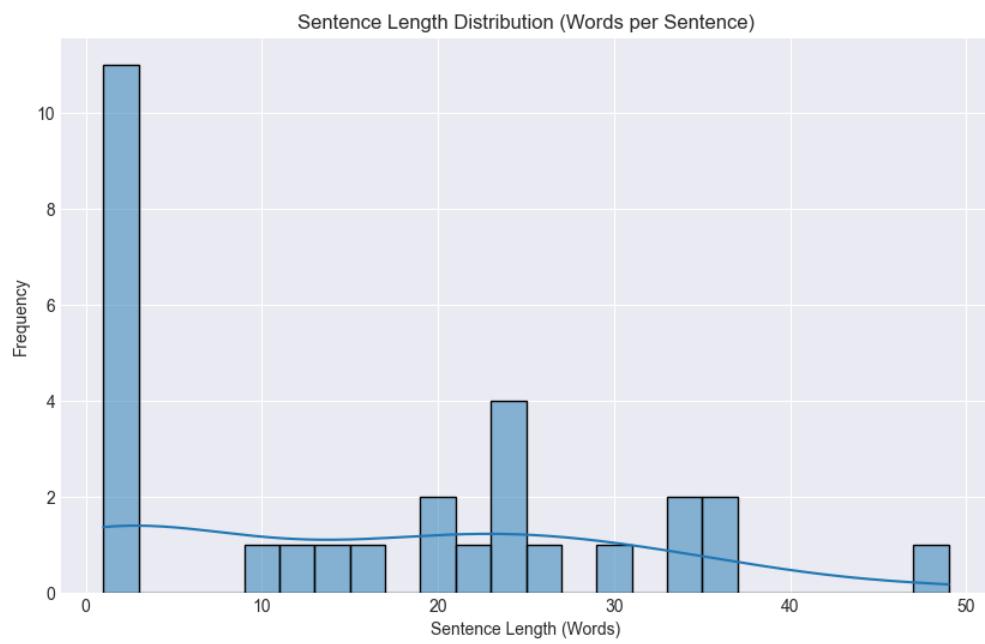
Index	Score
Flesch Reading Ease	38.25
Flesch-Kincaid Grade Level	14.00
Gunning Fog Index	14.50
SMOG Index	15.50
Coleman-Liau Index	14.10
Automated Readability Index	17.30

4. Visualizations

Word Length Distribution

Word Length Distribution

Sentence Length Distribution
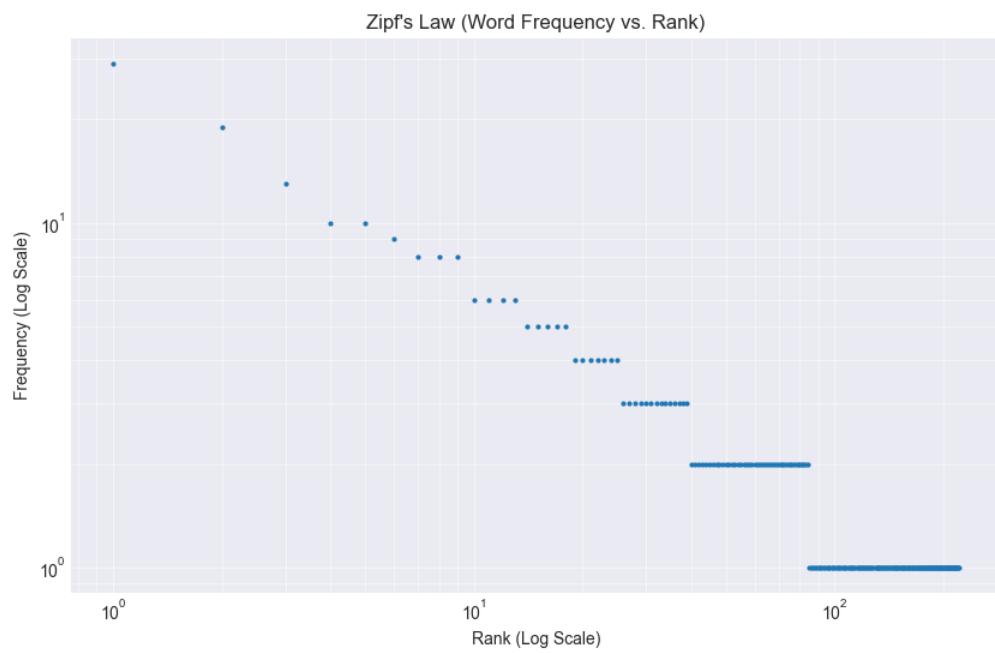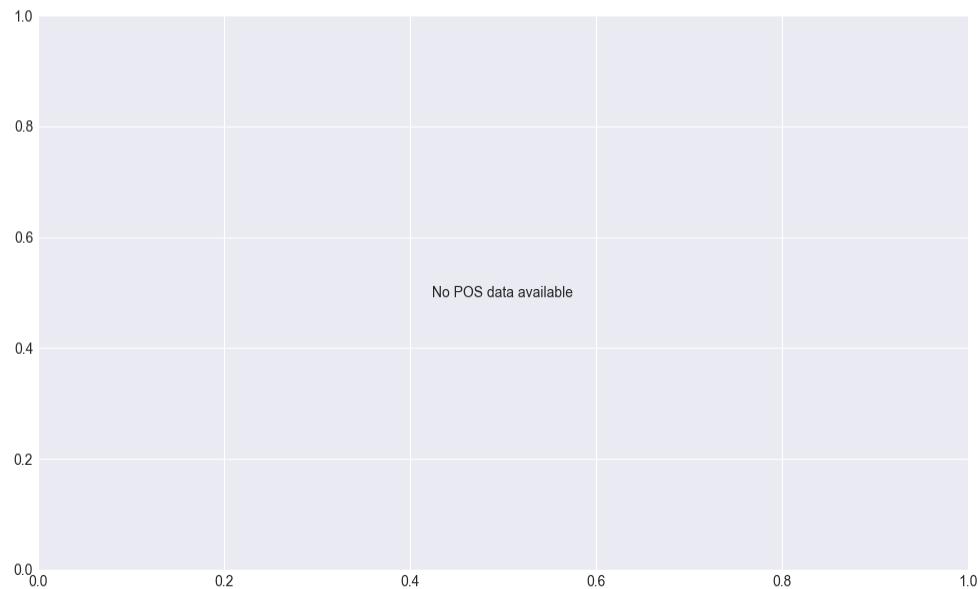
Sentence Length Distribution (Words per Sentence)
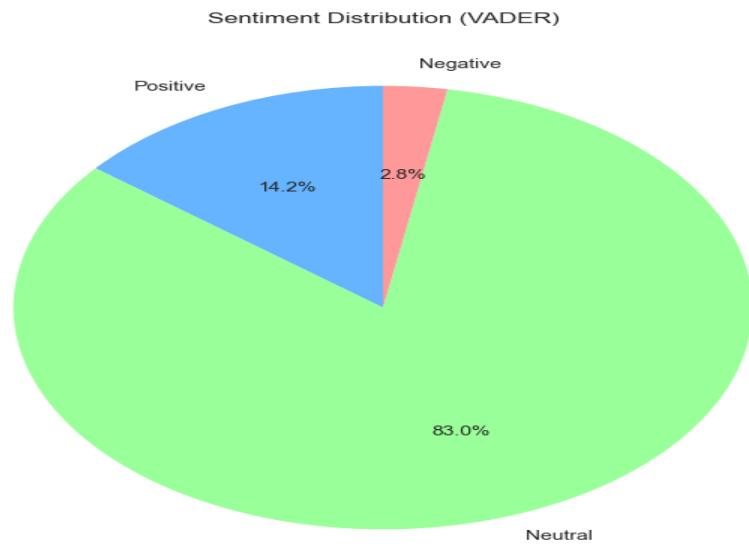
Zipf's Law Analysis (Word Frequency vs Rank)

Zipf's Law (Word Frequency vs. Rank)
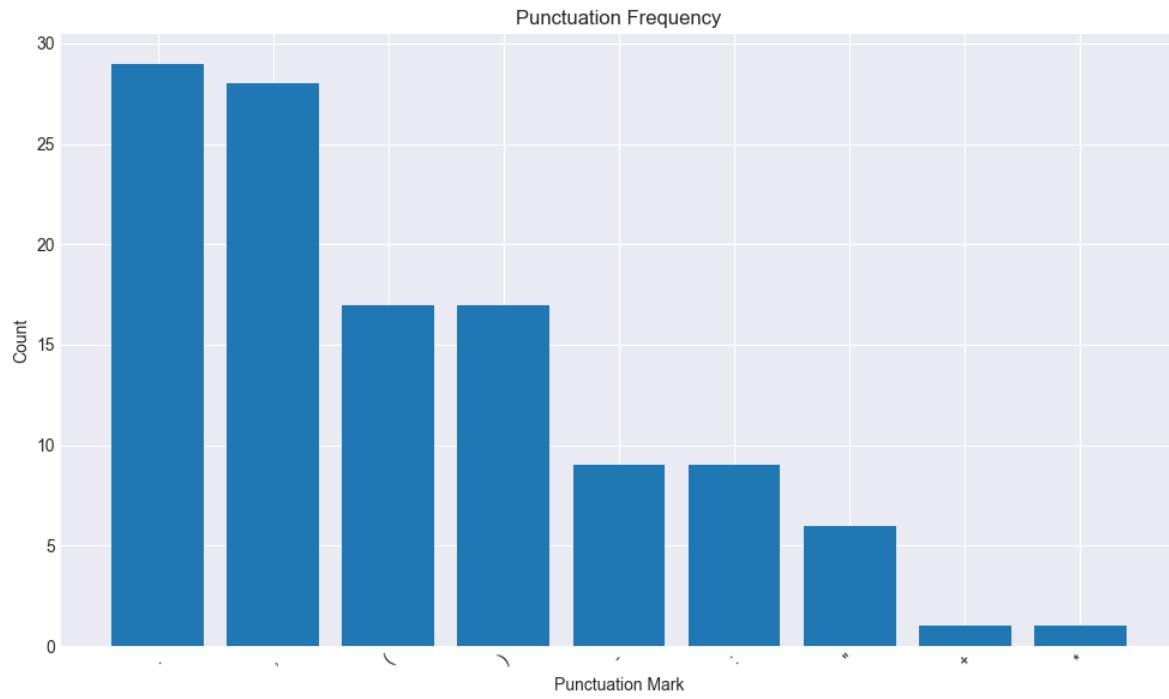
Part-of-Speech Distribution (Top 25)

No POS data available

Sentiment Distribution (VADER)

Sentiment Distribution (VADER)

Punctuation Frequency

5. Frequency Analysis

Most Common Words (Top 15, with stopwords):

the (29), and (19), with (13), marvel (10), in (10), television (9), of (8), phase (8), to (8), on (6), series (6), films (6), began (6), universe (5), by (5)

Most Common Punctuation:

'.' (29), ',' (28), '(' (17), ')' (17), '-' (9), ':' (9), '"' (6), '*' (1), '+' (1)

6. N-gram Analysis

Common 2-grams (Top 10):

'marvel studios' (4), 'by marvel' (3), 'the first' (3), 'as the' (3), 'and the' (3), 'began with' (3), 'concluded with' (3), 'shared universe' (2), 'a series' (2), 'series of' (2)
2-gram Repetition Rate: 0.1155

Common 3-grams (Top 10):

'a series of' (2), 'published by marvel' (2), 'by marvel comics' (2), 'and concluded with' (2), 'with ant man' (2), 'will conclude with' (2), 'agents of s' (2), 'of s h' (2), 's h i' (2), 'h i e' (2)
3-gram Repetition Rate: 0.0262

Common 4-grams (Top 10):

'published by marvel comics' (2), 'agents of s h' (2), 'of s h i' (2), 's h i e' (2), 'h i e l' (2), 'i e l d' (2), 'the marvel cinematic universe' (1), 'marvel cinematic universe mcu' (1), 'cinematic universe mcu is' (1), 'universe mcu is an' (1)
4-gram Repetition Rate: 0.0131

Character N-gram Analysis:

Common Character 3-grams (Top 10):

```
'_th' (36), 'the' (35), 'he_' (29), 'and' (24), '_an' (23), 'ed_' (23), '_20' (23), 'nd_' (20),
'se_' (19), 'on_' (19)
```

Common Character 4-grams (Top 10):

```
'_the' (30), 'the_' (29), 'and_' (20), '_and' (19), 'ing_' (16), '_wit' (13), 'with' (13),
'ith_' (13), '_201' (12), '_fil' (11)
```

Common Character 5-grams (Top 10):

```
'_the_' (27), '_and_' (19), '_with' (13), 'with_' (13), '_film' (11), '_phas' (11), 'phase'
(11), 'marve' (10), 'arvel' (10), 'rvel_' (10)
```