Text Analysis Report:
marvel_cinematic_universe_deepseek

1. Basic Statistics

Metric	Value
Word Count	448
Sentence Count	17
Character Count	2981
Unique Word Count	237
Avg Word Length (chars)	5.43
Avg Sentence Length (words)	26.35
Avg Sentence Length (chars)	175.35
Sentence Length Std Dev (words)	11.10
Lexical Diversity (TTR)	0.5290
MATTR (Window=50)	0.8265

## 2. Entropy & Complexity
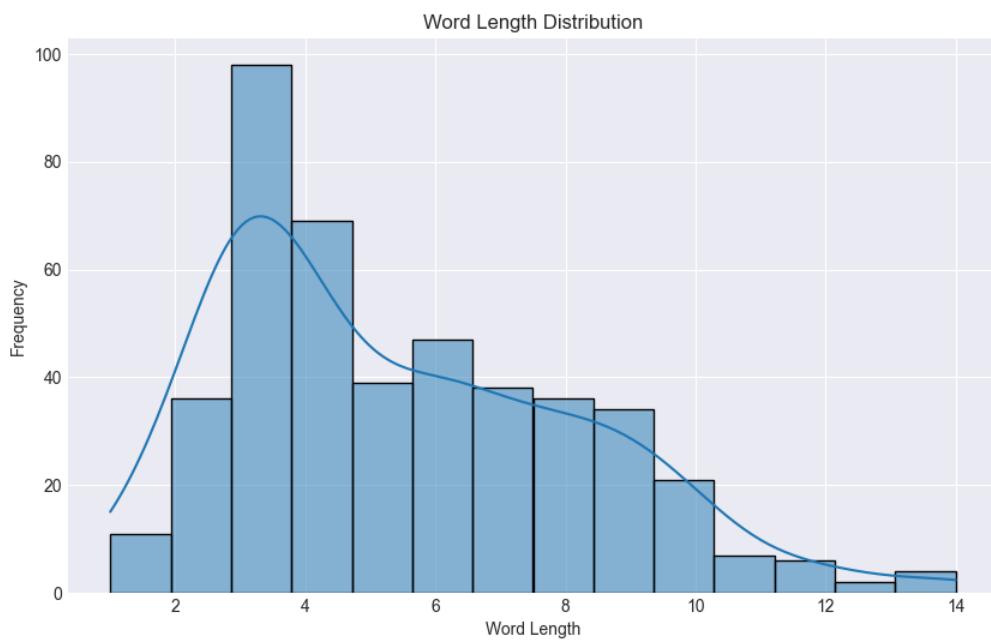
Metric	Value
Shannon Entropy (Words)	7.1888 bits
Shannon Entropy (Chars)	4.6295 bits
Function Word Ratio	0.3281

3. Readability Scores
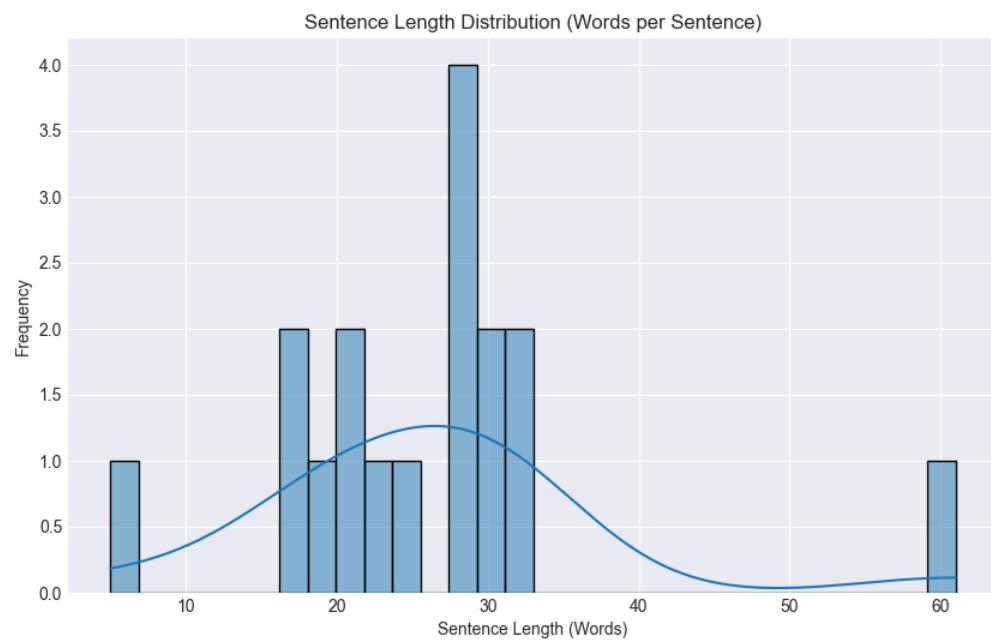
Index	Score
Flesch Reading Ease	28.67
Flesch-Kincaid Grade Level	15.60
Gunning Fog Index	16.67
SMOG Index	17.00
Coleman-Liau Index	15.55
Automated Readability Index	19.00

4. Visualizations

Word Length Distribution

Word Length Distribution

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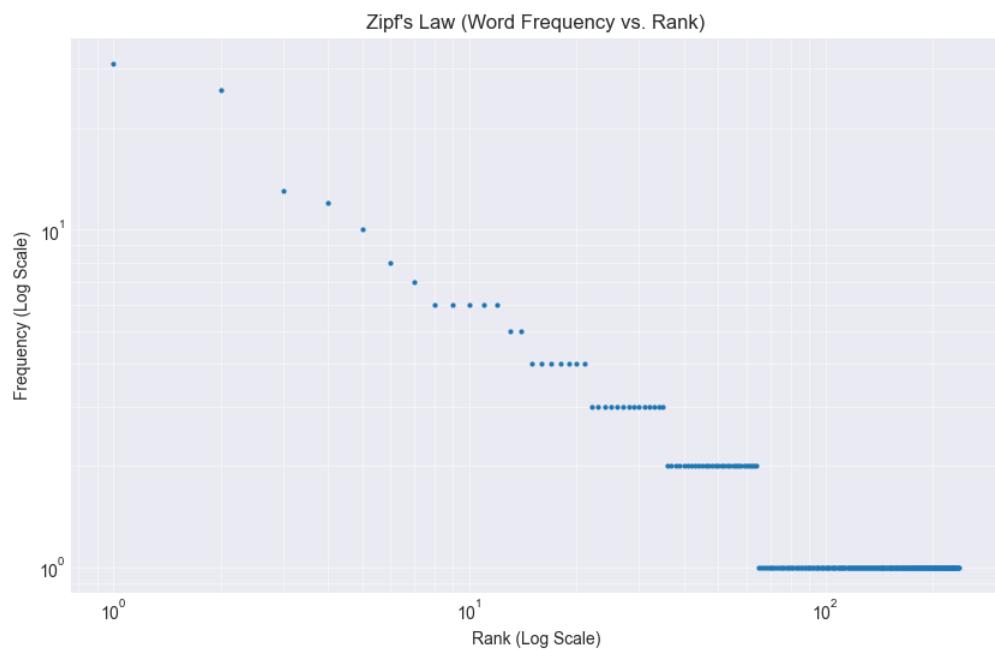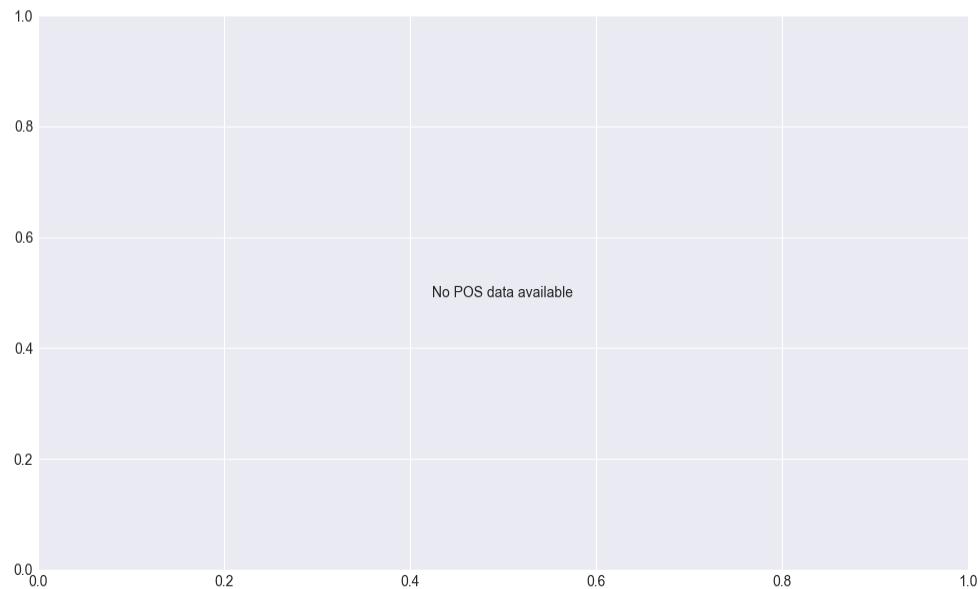
Sentence Length Distribution (Words per Sentence)
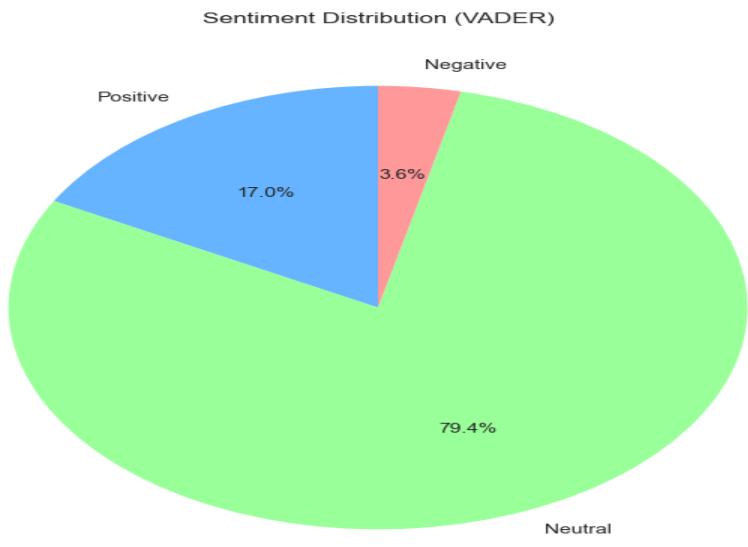
Zipf's Law Analysis (Word Frequency vs Rank)

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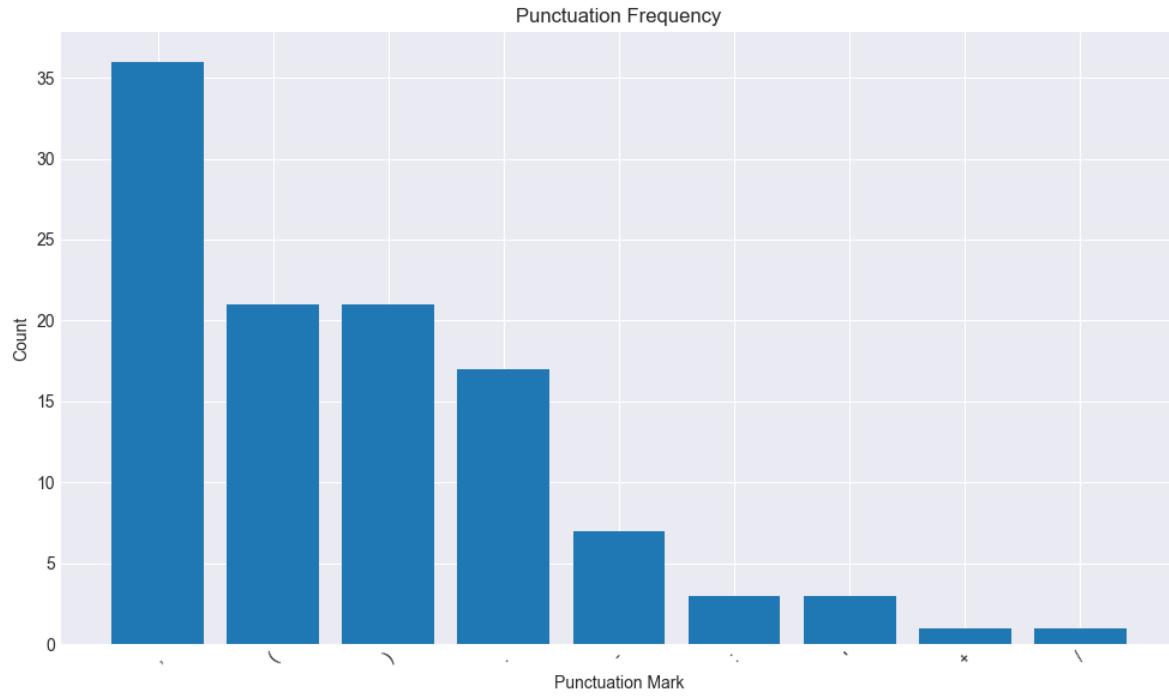
Part-of-Speech Distribution (Top 25)

No POS data available

Sentiment Distribution (VADER)

Sentiment Distribution (VADER)

Punctuation Frequency

Punctuation Frequency

5. Frequency Analysis

Most Common Words (Top 15, with stopwords):

the (31), and (26), of (13), with (12), mcu (10), a (8), its (7), marvel (6), universe (6), series (6), films (6), characters (6), in (5), like (5), franchise (4)

Most Common Punctuation:

',' (36), '(' (21), ')' (21), '.' (17), '-' (7), ':' (3), '"' (3), '/' (1), '+' (1)

6. N-gram Analysis

Common 2-grams (Top 10):

'the mcu' (8), 'films and' (4), 'marvel cinematic' (3), 'cinematic universe' (3), 'mcu is' (3), 'iron man' (3), 'with films' (3), 'mcu s' (3), 'universe mcu' (2), 'the marvel' (2)
2-gram Repetition Rate: 0.0984

Common 3-grams (Top 10):

'marvel cinematic universe' (3), 'the mcu s' (3), 'cinematic universe mcu' (2), 'the marvel cinematic' (2), 'a series of' (2), 'the mcu is' (2), 'the highest grossing' (2), 'highest grossing film' (2), 'of all time' (2), 'the mcu has' (2)
3-gram Repetition Rate: 0.0314

Common 4-grams (Top 10):

'marvel cinematic universe mcu' (2), 'the marvel cinematic universe' (2), 'the highest grossing film' (2), 'cinematic universe mcu the' (1), 'universe mcu the marvel' (1), 'mcu the marvel cinematic' (1), 'cinematic universe mcu is' (1), 'universe mcu is an' (1), 'mcu is an american' (1), 'is an american media' (1)
4-gram Repetition Rate: 0.0067

Character N-gram Analysis:

Common Character 3-grams (Top 10):

'the' (36), 'and' (33), '_th' (33), 'he_' (31), 'nd_' (31), '_an' (28), 'ng_' (26), 'ing' (25), 'es_' (22), 'ers' (21)

Common Character 4-grams (Top 10):

'the_' (31), '_the' (29), 'and_' (28), '_and' (26), 'ing_' (25), '_of_' (13), 'ers_' (12), '_wit' (12), 'with' (12), 'ith_' (12)

Common Character 5-grams (Top 10):

'_the_' (27), '_and_' (26), '_with' (12), 'with_' (12), 'the_m' (11), 'e_mcu' (10), '_film' (9), 's_and' (9), 'ivers' (8), 'verse' (8)