

Text Analysis Report: marvel_cinematic_universe_bard

1. Basic Statistics

Metric	Value
Word Count	296
Sentence Count	15
Character Count	2085
Unique Word Count	165
Avg Word Length (chars)	5.68
Avg Sentence Length (words)	19.73
Avg Sentence Length (chars)	139.00
Sentence Length Std Dev (words)	6.44
Lexical Diversity (TTR)	0.5574
MATTR (Window=50)	0.8301

2. Entropy & Complexity

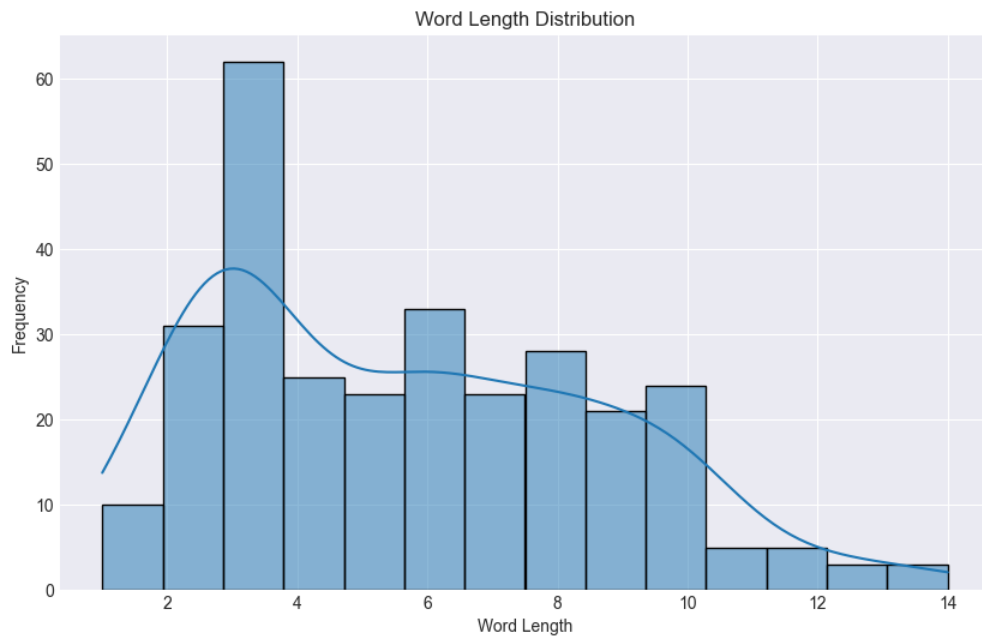
Metric	Value
Shannon Entropy (Words)	6.8426 bits
Shannon Entropy (Chars)	4.4991 bits
Function Word Ratio	0.3480

3. Readability Scores

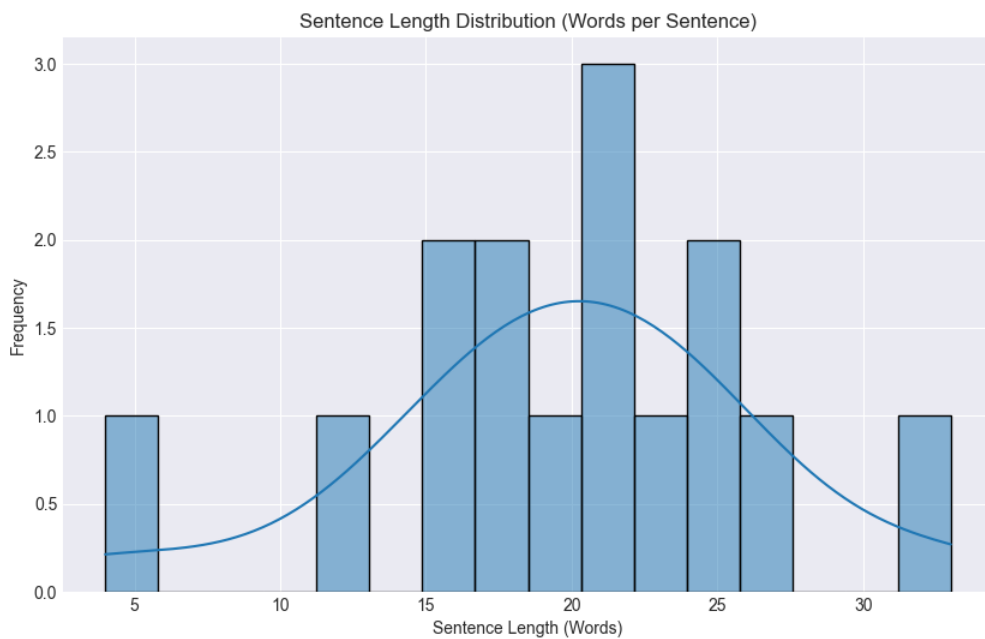
Index	Score
Flesch Reading Ease	26.30
Flesch-Kincaid Grade Level	14.40
Gunning Fog Index	15.61
SMOG Index	16.60
Coleman-Liau Index	16.13
Automated Readability Index	16.50

4. Visualizations

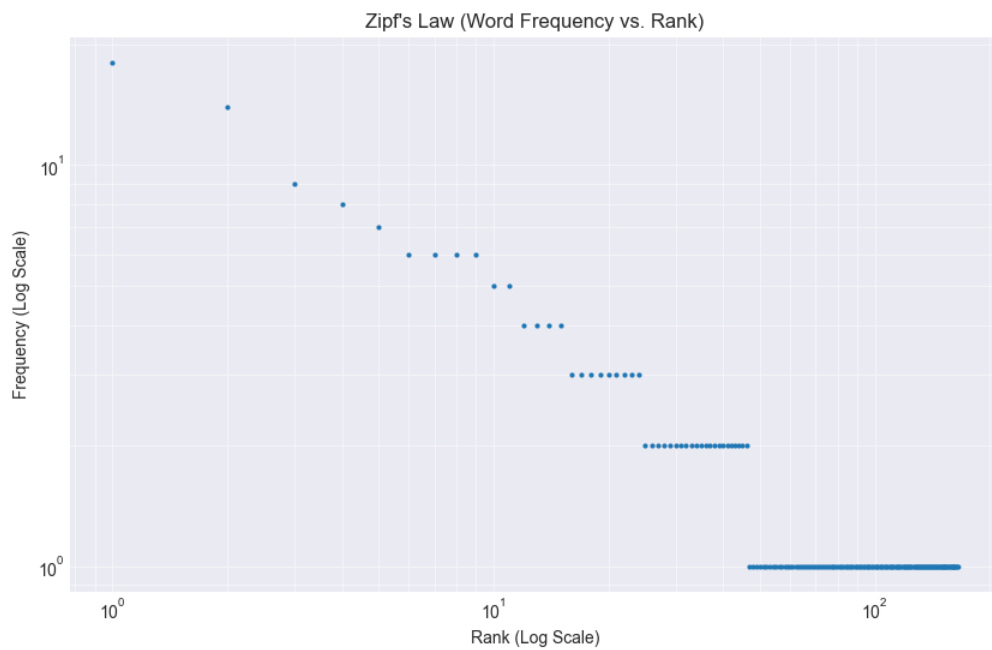
Word Length Distribution

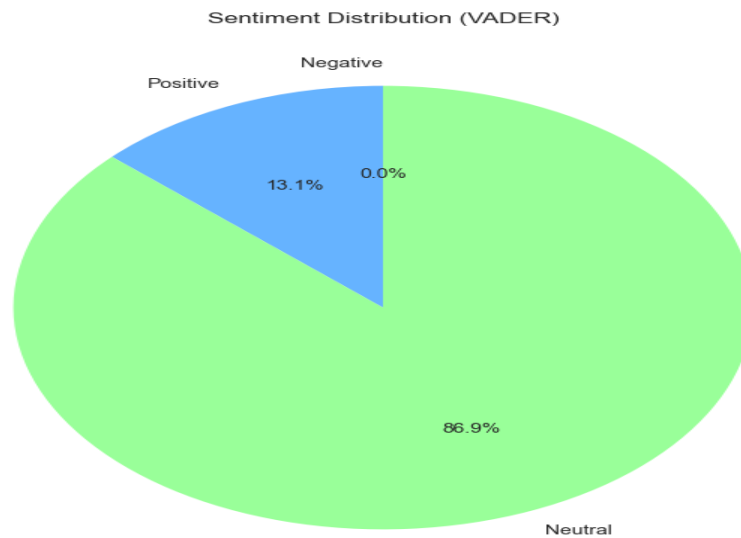


Sentence Length Distribution

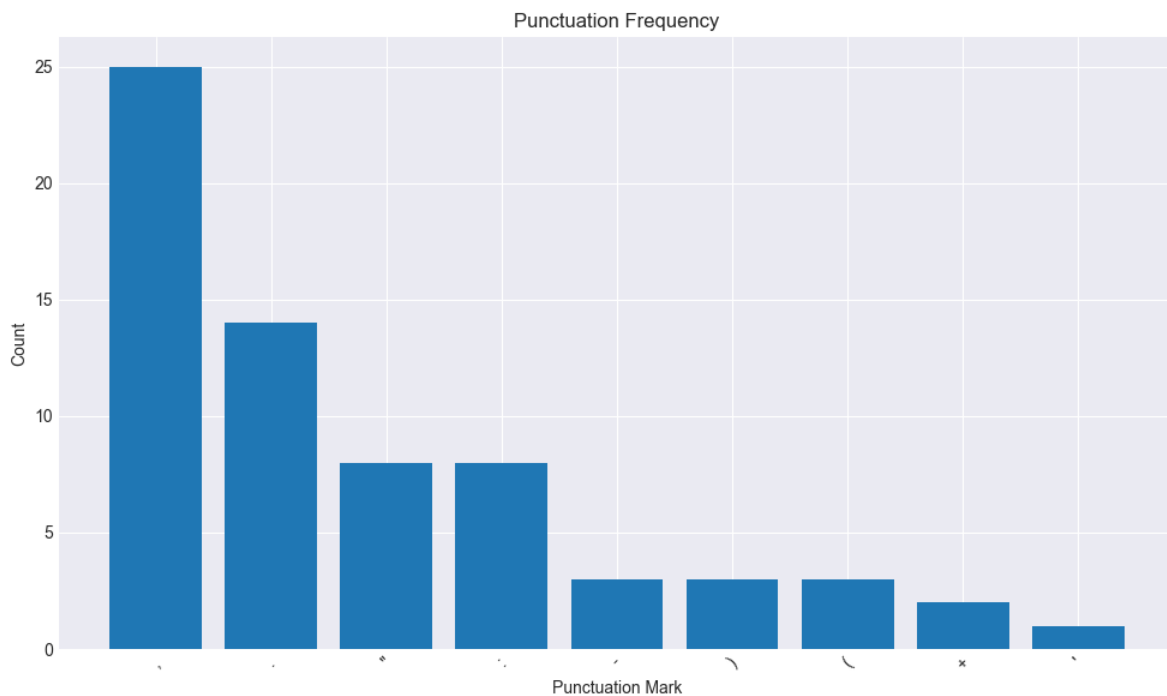


Zipf's Law Analysis (Word Frequency vs Rank)





Punctuation Frequency



5. Frequency Analysis

Most Common Words (Top 15, with stopwords):

the (18), and (14), a (9), of (8), mcu (7), universe (6), films (6), in (6), has (6), series (5), television (5), is (4), shared (4), it (4), phases (4)

Most Common Punctuation:

',' (25), ':' (14), ':' (8), '"' (8), '(' (3), ')' (3), '-' (3), '+' (2), "'" (1)

6. N-gram Analysis

Common 2-grams (Top 10):

'the mcu' (6), 'shared universe' (4), 'films and' (3), 'television series' (3), 'it has' (3), 'of the' (3), 'marvel studios' (2), 'and television' (2), 'creating a' (2), 'with the' (2)
2-gram Repetition Rate: 0.0780

Common 3-grams (Top 10):

'the mcu has' (2), 'the marvel cinematic' (1), 'marvel cinematic universe' (1), 'cinematic universe mcu' (1), 'universe mcu is' (1), 'mcu is an' (1), 'is an american' (1), 'an american media' (1), 'american media franchise' (1), 'media franchise and' (1)
3-gram Repetition Rate: 0.0034

Common 4-grams (Top 10):

'the marvel cinematic universe' (1), 'marvel cinematic universe mcu' (1), 'cinematic universe mcu is' (1), 'universe mcu is an' (1), 'mcu is an american' (1), 'is an american media' (1), 'an american media franchise' (1), 'american media franchise and' (1), 'media franchise and shared' (1), 'franchise and shared universe' (1)
4-gram Repetition Rate: 0.0000

Character N-gram Analysis:

Common Character 3-grams (Top 10):

'the' (20), 'es_' (20), 'he_' (18), 'ed_' (18), 'ing' (18), 'ng_' (18), 'and' (17), '_th' (17), '_an' (15), 'nd_' (15)

Common Character 4-grams (Top 10):

'the_' (18), 'ing_' (18), '_and' (14), 'and_' (14), '_ _ ' (13), '_the' (13), 'he_m' (10), 'ion_' (9), 'iver' (8), 'vers' (8)

Common Character 5-grams (Top 10):

'_and_' (14), '_the_' (13), 'the_m' (10), 'ivers' (8), 'verse' (8), 'erse_' (8), '_film' (8), 'e_mcu' (7), '_mcu_' (7), 's_and' (7)