

Text Analysis Report: fall_of_roman_empire_deepseek

1. Basic Statistics

Metric	Value
Word Count	502
Sentence Count	18
Character Count	3086
Unique Word Count	251
Avg Word Length (chars)	5.00
Avg Sentence Length (words)	27.89
Avg Sentence Length (chars)	171.44
Sentence Length Std Dev (words)	9.33
Lexical Diversity (TTR)	0.5000
MATTR (Window=50)	0.7571

2. Entropy & Complexity

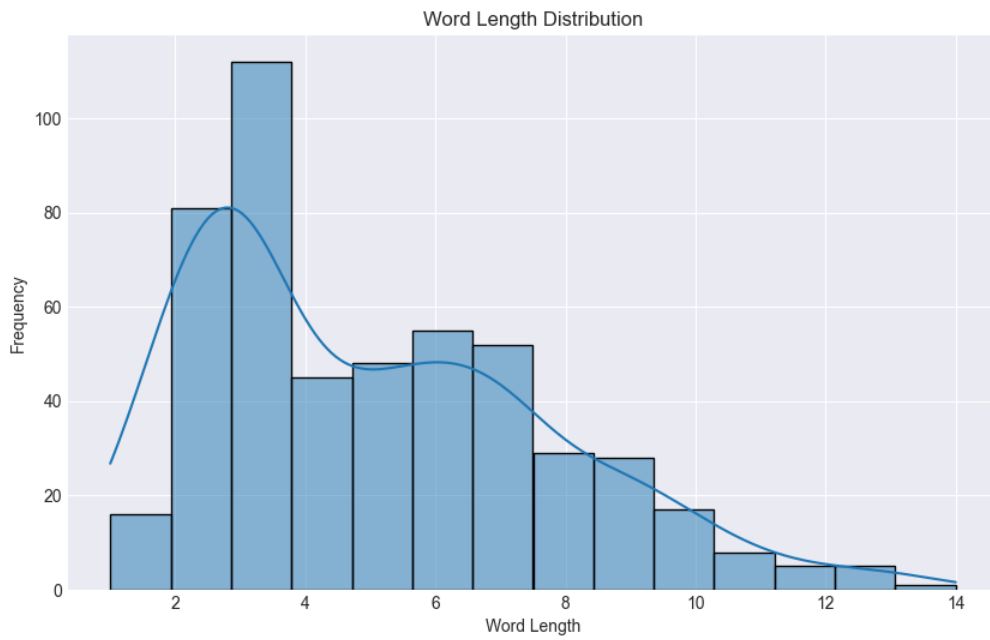
Metric	Value
Shannon Entropy (Words)	6.8459 bits
Shannon Entropy (Chars)	4.4507 bits
Function Word Ratio	0.4203

3. Readability Scores

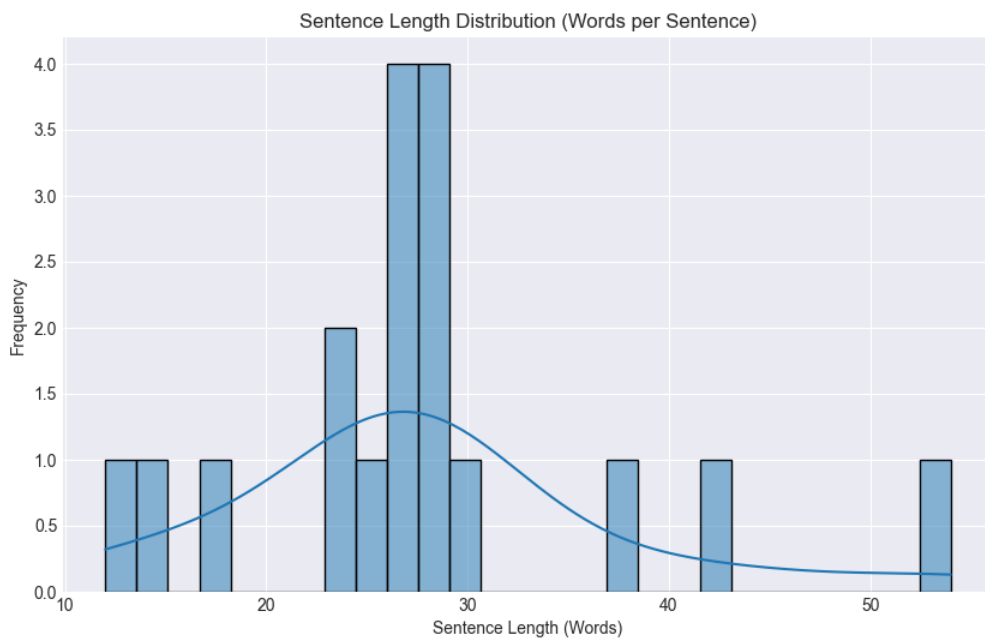
Index	Score
Flesch Reading Ease	26.54
Flesch-Kincaid Grade Level	16.40
Gunning Fog Index	17.41
SMOG Index	17.60
Coleman-Liau Index	12.36
Automated Readability Index	16.90

4. Visualizations

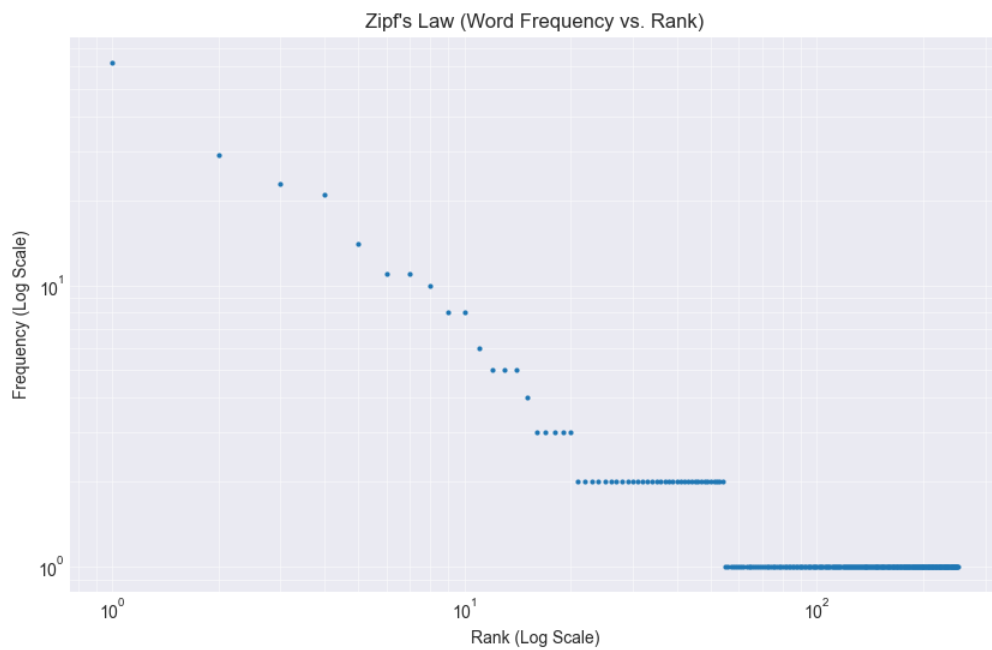
Word Length Distribution



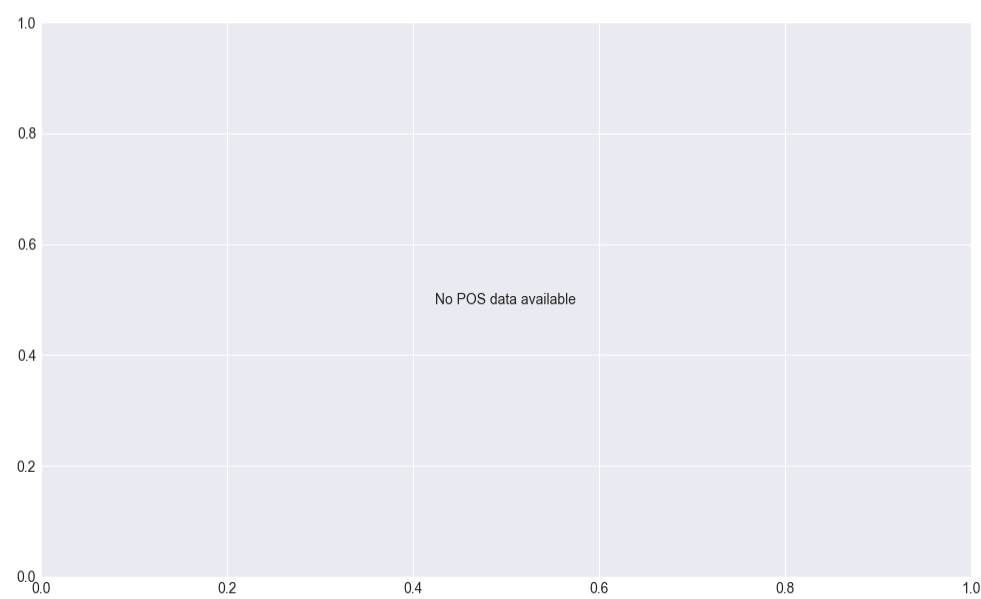
Sentence Length Distribution



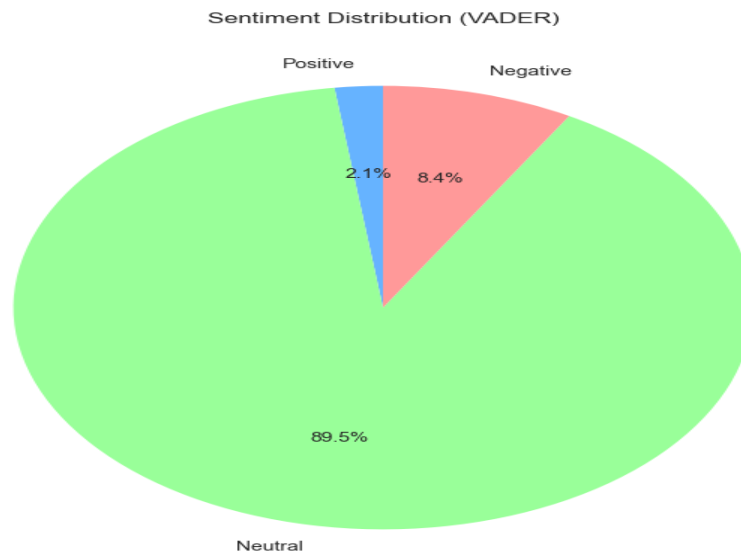
Zipf's Law Analysis (Word Frequency vs Rank)



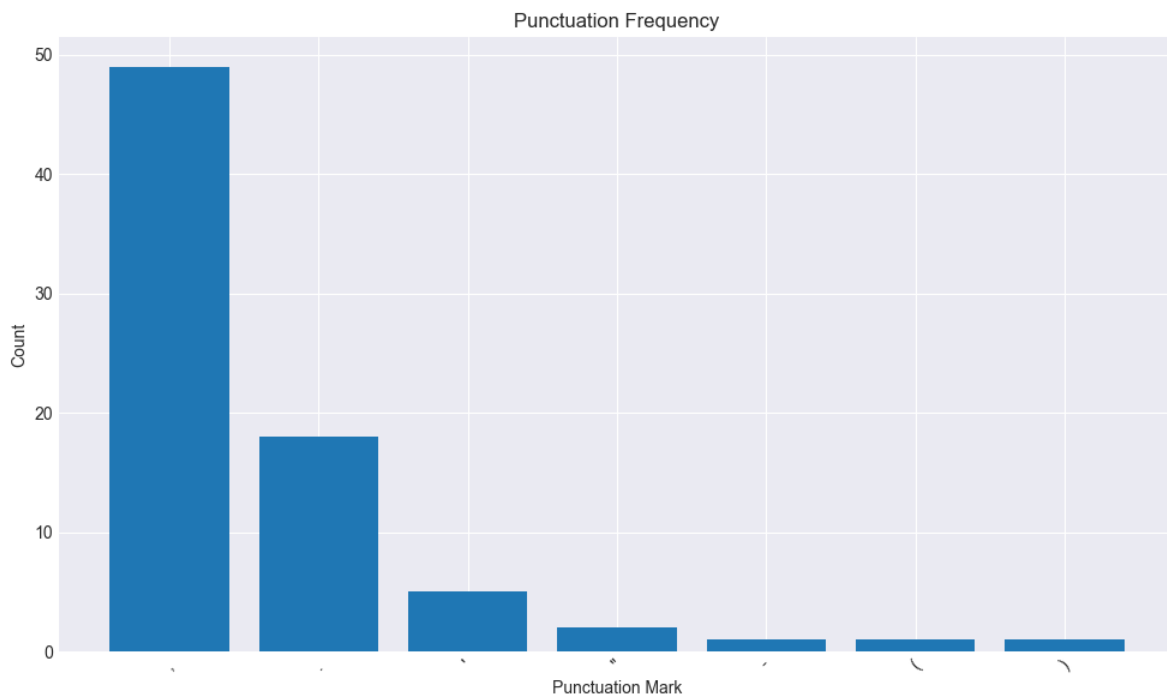
Part-of-Speech Distribution (Top 25)



Sentiment Distribution (VADER)



Punctuation Frequency



5. Frequency Analysis

Most Common Words (Top 15, with stopwords):

the (62), of (29), and (23), empire (21), roman (14), a (11), to (11), in (10), by (8), as (8), ce (6), fall (5), western (5), s (5), decline (4)

Most Common Punctuation:

',' (49), '.' (18), '"' (5), "'" (2), '(' (1), ')' (1), '-' (1)

6. N-gram Analysis

Common 2-grams (Top 10):

'of the' (12), 'roman empire' (10), 'the empire' (9), 'and the' (8), 'the roman' (7), 'to the' (5), 'empire s' (5), 'fall of' (4), 'the fall' (4), 'western roman' (4)

2-gram Repetition Rate: 0.1756

Common 3-grams (Top 10):

'of the roman' (5), 'the roman empire' (5), 'the empire s' (5), 'fall of the' (3), 'the fall of' (3), 'the decline of' (3), 'the western roman' (3), 'western roman empire' (3), 'decline of the' (2), 'the eastern roman' (2)

3-gram Repetition Rate: 0.0560

Common 4-grams (Top 10):

'of the roman empire' (4), 'fall of the roman' (3), 'the western roman empire' (3), 'the fall of the' (2), 'the decline of the' (2), 'the eastern roman empire' (2), 'of the western roman' (2), 'the roman empire the' (1), 'roman empire the fall' (1), 'empire the fall of' (1)

4-gram Repetition Rate: 0.0220

Character N-gram Analysis:

Common Character 3-grams (Top 10):

'the' (65), '_th' (64), 'he_' (62), '_of' (31), 'of_' (29), 'nd_' (27), '_an' (25), 'and' (24), '_em' (23), 'emp' (23)

Common Character 4-grams (Top 10):

'the_' (62), '_the' (60), '_of_' (29), '_emp' (23), '_and' (23), 'and_' (23), 'empi' (21), 'mpir' (21), 'pire' (21), '_rom' (18)

Common Character 5-grams (Top 10):

'_the_' (59), '_and_' (23), '_empi' (21), 'empir' (21), 'mpire' (21), 'pire_' (15), '_roma' (14), 'roman' (14), 'oman_' (14), '_of_t' (13)