

Text Analysis Report: fall_of_roman_empire_wikipedia

1. Basic Statistics

Metric	Value
Word Count	457
Sentence Count	18
Character Count	2886
Unique Word Count	233
Avg Word Length (chars)	5.18
Avg Sentence Length (words)	25.39
Avg Sentence Length (chars)	160.33
Sentence Length Std Dev (words)	14.61
Lexical Diversity (TTR)	0.5098
MATTR (Window=50)	0.7796

2. Entropy & Complexity

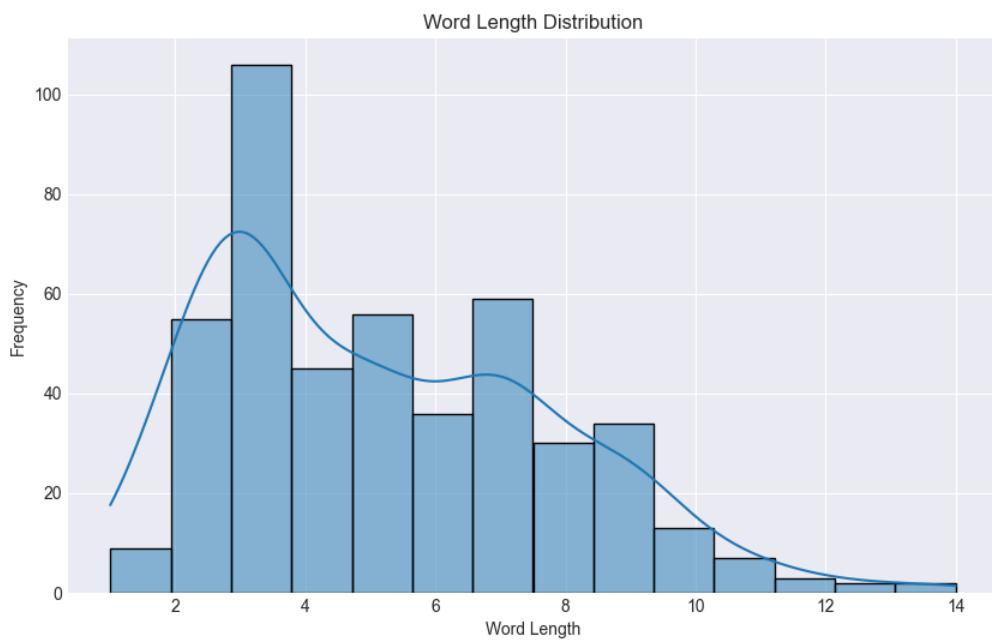
Metric	Value
Shannon Entropy (Words)	6.9143 bits
Shannon Entropy (Chars)	4.3756 bits
Function Word Ratio	0.4114

3. Readability Scores

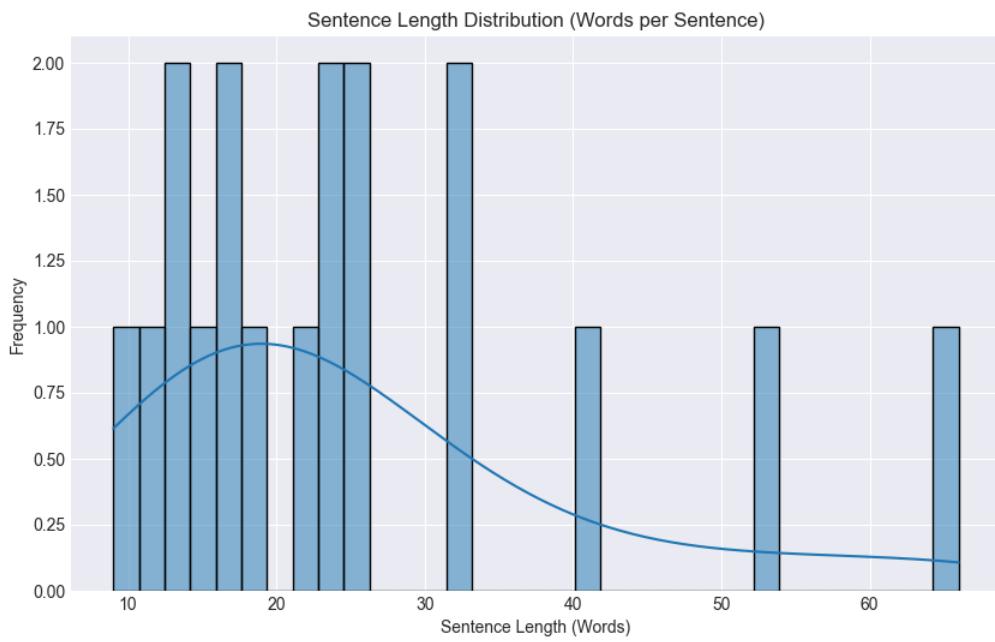
Index	Score
Flesch Reading Ease	29.08
Flesch-Kincaid Grade Level	15.40
Gunning Fog Index	16.07
SMOG Index	16.30
Coleman-Liau Index	13.47
Automated Readability Index	16.50

4. Visualizations

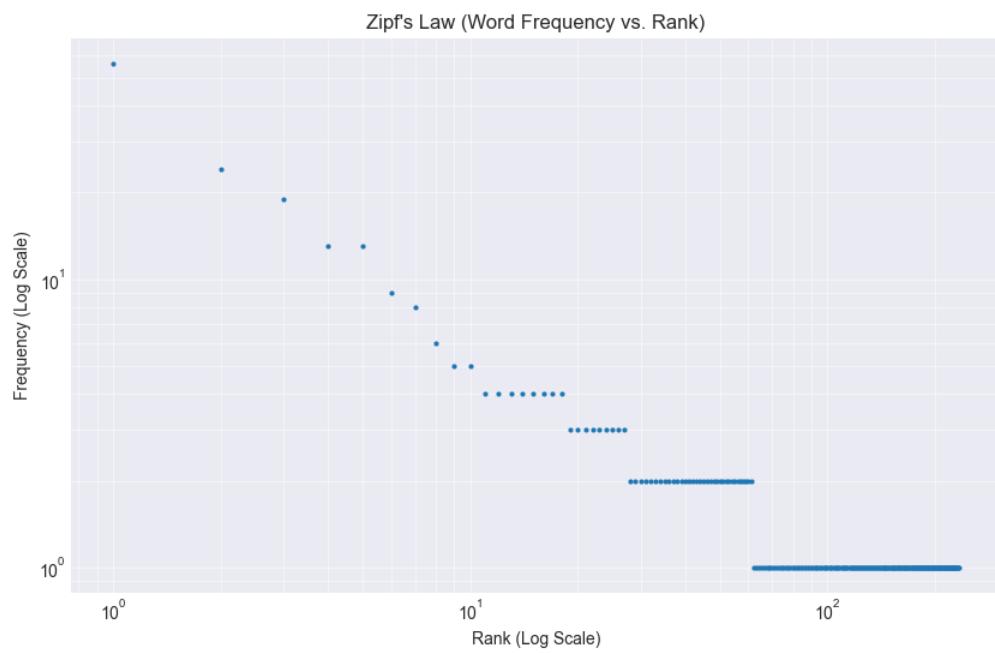
Word Length Distribution



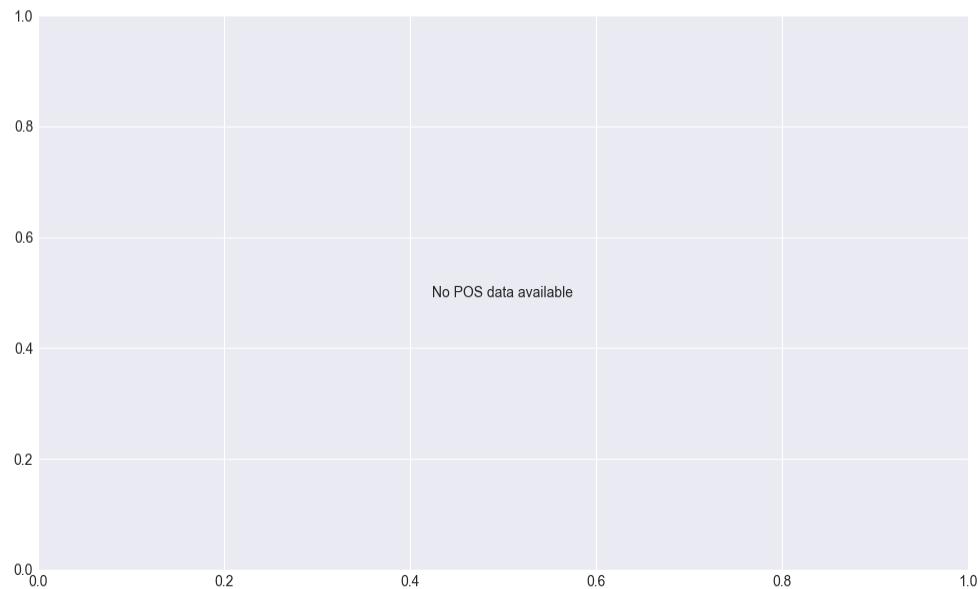
Sentence Length Distribution



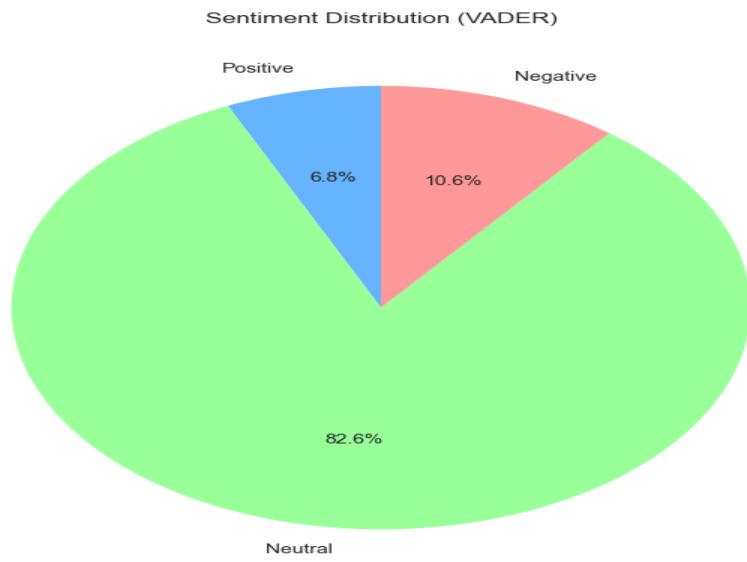
Zipf's Law Analysis (Word Frequency vs Rank)



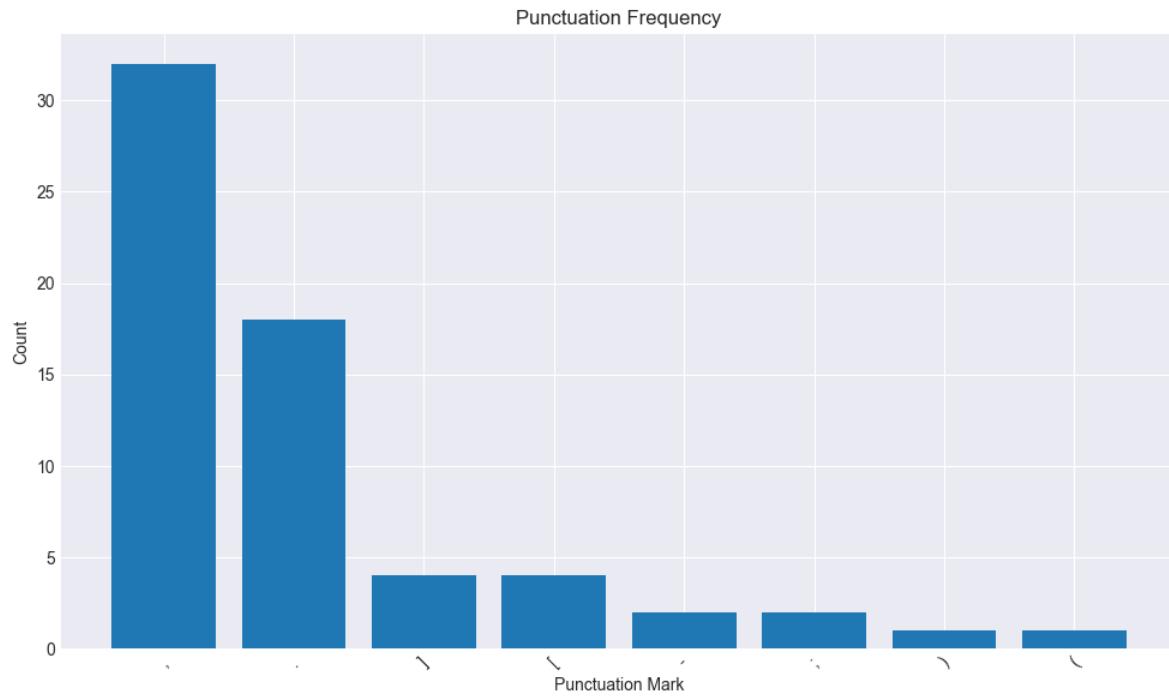
Part-of-Speech Distribution (Top 25)



Sentiment Distribution (VADER)



Punctuation Frequency



5. Frequency Analysis

Most Common Words (Top 15, with stopwords):

the (56), of (24), and (19), roman (13), empire (13), western (9), in (8), to (6), its (5), for (5), fall (4), or (4), was (4), political (4), control (4)

Most Common Punctuation:

',' (32), '.' (18), '[' (4), ']' (4), ';' (2), '-' (2), '(' (1), ')' (1)

6. N-gram Analysis

Common 2-grams (Top 10):

'of the' (16), 'the western' (6), 'roman empire' (5), 'the fall' (4), 'fall of' (4), 'western roman' (4), 'the empire' (4), 'the roman' (3), 'and the' (3), 'and other' (3)

2-gram Repetition Rate: 0.1469

Common 3-grams (Top 10):

'the fall of' (4), 'of the western' (4), 'the western roman' (3), 'western roman empire' (3), 'the western empire' (3), 'fall of the' (2), 'of the roman' (2), 'the roman empire' (2), 'fall of rome' (2), 'the loss of' (2)

3-gram Repetition Rate: 0.0505

Common 4-grams (Top 10):

'the western roman empire' (3), 'the fall of the' (2), 'of the western roman' (2), 'the fall of rome' (2), 'and numbers of the' (2), 'goths and other non' (2), 'of the western empire' (2), 'fall of the western' (1), 'western roman empire also' (1), 'roman empire also called' (1)

4-gram Repetition Rate: 0.0176

Character N-gram Analysis:

Common Character 3-grams (Top 10):

'the' (68), '_th' (66), 'he_' (57), '_of' (24), 'of_' (24), '_an' (22), 'ed_' (21), 'ter' (20), 'nd_' (20), 'rom' (19)

Common Character 4-grams (Top 10):

'_the' (62), 'the_' (56), '_of_' (24), '_and' (19), 'and_' (19), '_emp' (18), 'of_t' (17), 'f_th' (17), '_rom' (15), 'ern_' (14)

Common Character 5-grams (Top 10):

'_the_' (55), '_and_' (19), '_of_t' (17), 'of_th' (17), 'f_the' (17), 'roman' (14), 'oman_' (13), '_empi' (13), 'empir' (13), 'mpire' (13)