

Text Analysis Report: blockchain_technology_wikipedia

1. Basic Statistics

Metric	Value
Word Count	355
Sentence Count	14
Character Count	2373
Unique Word Count	211
Avg Word Length (chars)	5.50
Avg Sentence Length (words)	25.36
Avg Sentence Length (chars)	169.50
Sentence Length Std Dev (words)	10.84
Lexical Diversity (TTR)	0.5944
MATTR (Window=50)	0.8537

2. Entropy & Complexity

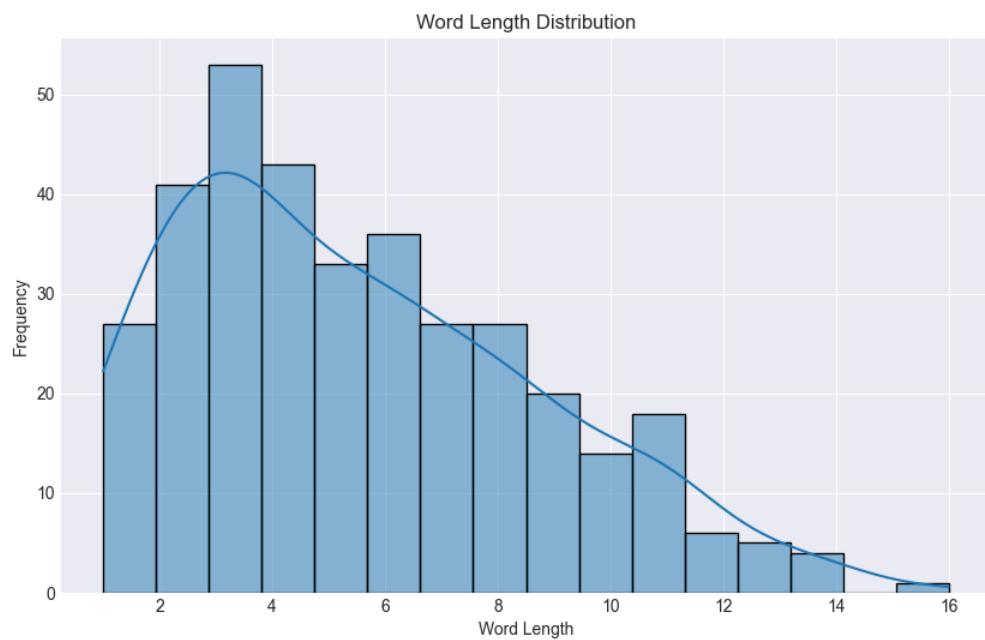
Metric	Value
Shannon Entropy (Words)	7.2409 bits
Shannon Entropy (Chars)	4.4736 bits
Function Word Ratio	0.3549

3. Readability Scores

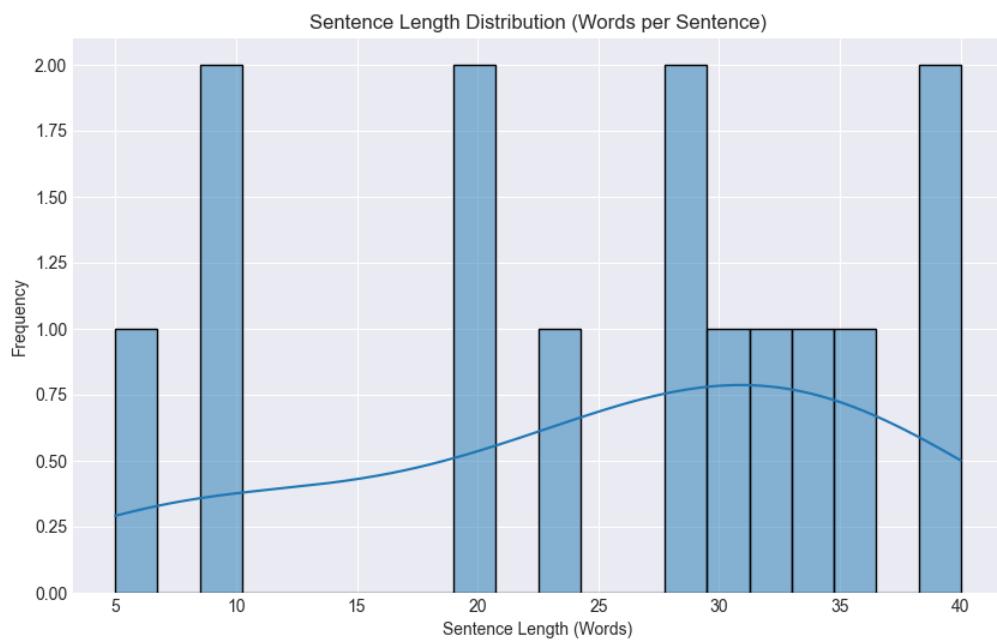
Index	Score
Flesch Reading Ease	29.89
Flesch-Kincaid Grade Level	15.10
Gunning Fog Index	16.66
SMOG Index	16.40
Coleman-Liau Index	16.31
Automated Readability Index	18.80

4. Visualizations

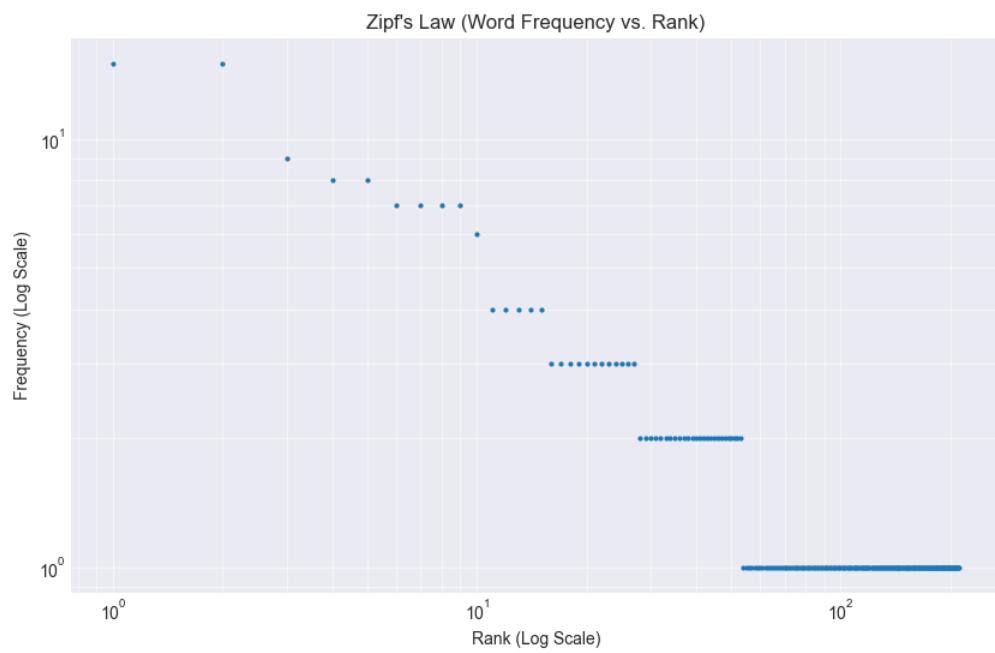
Word Length Distribution



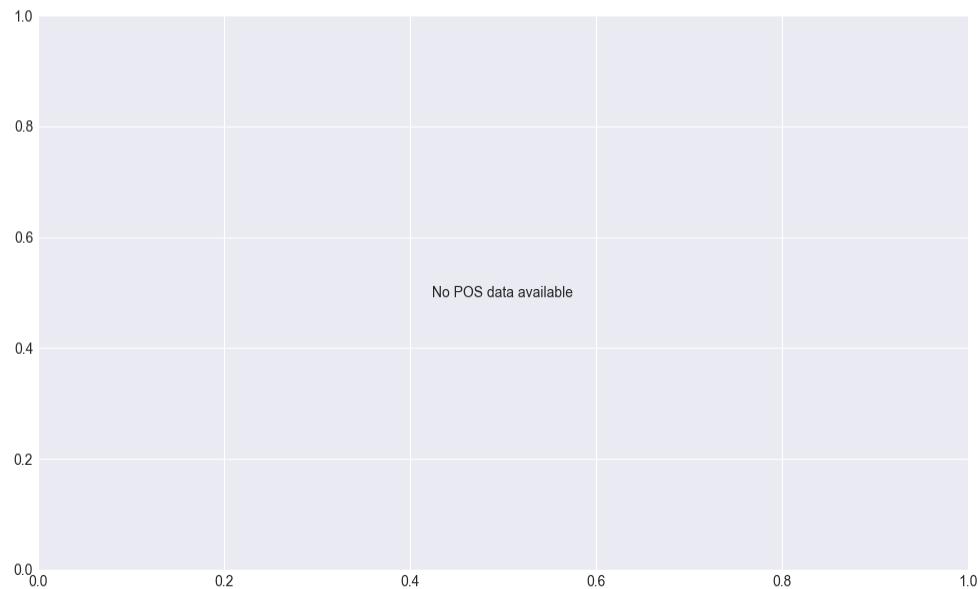
Sentence Length Distribution



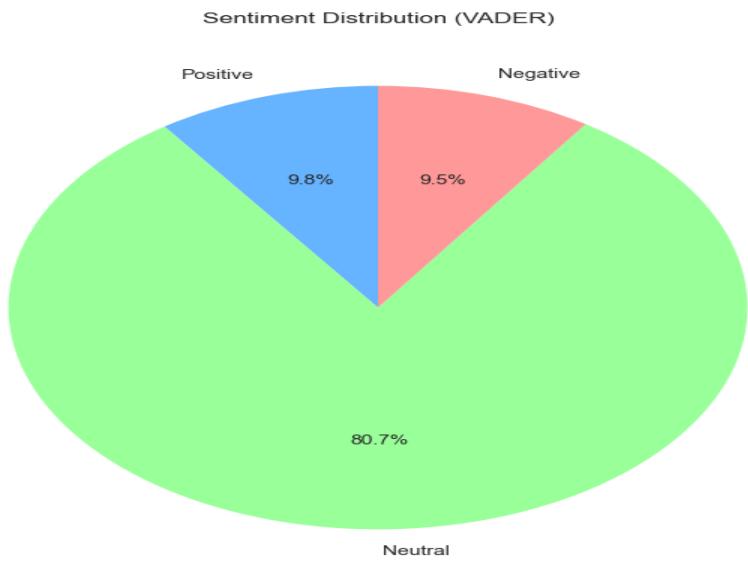
Zipf's Law Analysis (Word Frequency vs Rank)



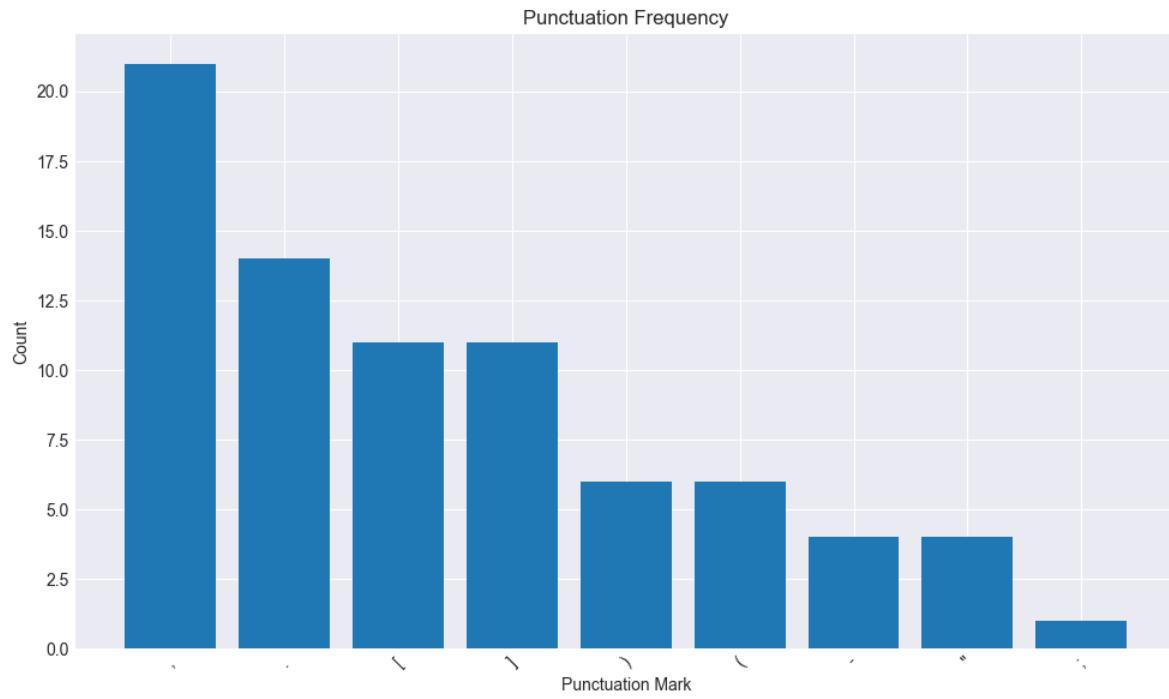
Part-of-Speech Distribution (Top 25)



Sentiment Distribution (VADER)



Punctuation Frequency



5. Frequency Analysis

Most Common Words (Top 15, with stopwords):

a (15), the (15), and (9), are (8), to (8), blockchain (7), of (7), by (7), blockchains (7), block (6), distributed (4), data (4), as (4), be (4), for (4)

Most Common Punctuation:

',' (21), '.' (14), '[' (11), ']' (11), '(' (6), ')' (6), """ (4), '-' (4), ';' (1)

6. N-gram Analysis

Common 2-grams (Top 10):

'distributed ledger' (3), 'may be' (3), 'a blockchain' (2), 'a distributed' (2), 'that are' (2), 'each block' (2), 'block contains' (2), 'of the' (2), 'the previous' (2), 'previous block' (2)

2-gram Repetition Rate: 0.0537

Common 3-grams (Top 10):

'each block contains' (2), 'the previous block' (2), 'public distributed ledger' (2), 'may be considered' (2), 'a blockchain is' (1), 'blockchain is a' (1), 'is a distributed' (1), 'a distributed ledger' (1), 'distributed ledger with' (1), 'ledger with growing' (1)

3-gram Repetition Rate: 0.0113

Common 4-grams (Top 10):

'a blockchain is a' (1), 'blockchain is a distributed' (1), 'is a distributed ledger' (1), 'a distributed ledger with' (1), 'distributed ledger with growing' (1), 'ledger with growing lists' (1), 'with growing lists of' (1), 'growing lists of records' (1), 'lists of records blocks' (1), 'of records blocks that' (1)

4-gram Repetition Rate: 0.0000

Character N-gram Analysis:

Common Character 3-grams (Top 10):

'ed_' (26), '_th' (25), 'blo' (23), 'loc' (23), 'ock' (23), '_bl' (22), 'the' (22), 'ain' (19), 're_' (19), 'cha' (17)

Common Character 4-grams (Top 10):

'bloc' (23), 'lock' (23), '_blo' (22), '_the' (19), 'chai' (15), 'hain' (15), 'the_' (15), 'ockc' (14), 'ckch' (14), 'kcha' (14)

Common Character 5-grams (Top 10):

'block' (23), '_bloc' (22), 'chain' (15), '_the_' (15), 'lockc' (14), 'ockch' (14), 'ckcha' (14), 'kchai' (14), 'ains_' (9), '_and_' (9)