BusinessDay Article Analysis Report

1. Project Overview

This research project involved scraping, processing, and analyzing business articles from BusinessDay.ng under the Business & Economy section. The objective was to extract valuable insights using Natural Language Processing (NLP) and visualize patterns in sentiment, topics, and authorship.

2. Tools & Libraries Used

- Programming Language: Python
- Web Scraping: requests, BeautifulSoup, Selenium
- Data Handling: pandas, numpy
- NLP: nltk, rake-nltk, vaderSentiment
- Visualization: matplotlib, seaborn, wordcloud

3. Methodology

a. Data Collection (Scraping):

- Extracted article URLs and parsed individual pages to retrieve title, author, date, and content.

b. Data Cleaning & Preprocessing:

- Tokenization, Lowercasing, Removing Punctuation and Stopwords, Lemmatization.

c. Text Analysis:

- Sentiment analysis using VADER and keyword extraction using RAKE.

4. Key Findings

- Sentiment Distribution: Majority of articles were neutral, followed by positive.
- Topic Insights: Common keywords were inflation, CBN, exchange rate, economic growth.
- Authorship: Most articles lacked author metadata due to extraction challenges.

5. Challenges Encountered

- Missing Author Info: Resolved with fallback logic.
- 403 Errors: Bypassed using headers, retries, and Selenium with undetected_chromedriver.

6. Conclusion

This project demonstrated the application of web scraping and NLP for media content analysis. Key insights from sentiment and keyword trends provide valuable understanding of business discourse. Future improvements could include time-based sentiment tracking and dashboard integration.