

Business Day News Article Analysis Report

Category: Economy | Source: BusinessDay.ng

1. Project Overview

This project involves scraping and analyzing articles from the 'Economy' category on the Business Day Nigeria website. The main objective is to extract article data, perform sentiment analysis, summarize content, and identify key insights through text analytics and visualizations.

2. Tools & Technologies Used

- Web Scraping: Octoparse, requests, BeautifulSoup
- Data Handling: pandas, datetime, os
- Text Preprocessing: nltk, re, string
- Sentiment Analysis: nltk.sentiment.vader
- Keyword Extraction: rake_nltk
- Visualization: matplotlib, seaborn, wordcloud

3. Methodology

Step 1: Data Collection

- Octoparse was used to scrape URLs from the Business Day 'Economy' category.

Step 2: Web Scraping

- Scraped Title, Author, Date, Full Content using requests + BeautifulSoup.

Step 3: Sentiment Analysis

- Applied VADER to extract sentiment and polarity scores.

Step 4: Content Summarization and Keyword Extraction

- Used NLTK and RAKE.

Step 5: Text Preprocessing

- Cleaned and lemmatized text.

Step 6: Re-running Sentiment Analysis

- Measured effect of preprocessing on sentiment.

Business Day News Article Analysis Report

Category: Economy | Source: BusinessDay.ng

4. Challenges & Resolutions

- 403 errors: Resolved with user-agent headers.
- Missing author: Used `.get()` or `try/except`.
- Sentiment drift post-cleaning: Kept raw + cleaned scores.
- Noisy keywords: RAKE moved to cleaned content.
- Inconsistent date format: Parsed with `pd.to_datetime(errors='coerce')`.

5. Exploratory Analysis & Visualizations

- Positive sentiment dominated.
- Articles with highest positivity clustered around policy wins or trade developments.
- Polarity dipped occasionally, marking likely economic or policy setbacks.
- Frequent authors and keywords indicated editorial priorities.

6. Key Findings

- Most articles were positive in tone.
- Text preprocessing altered polarity in some cases.
- Top keywords: CBN, inflation, investment, FX, budget.
- Journalism had recurring authors and themes over time.

7. Output Files

- Excel with article metadata
- CSV with full analysis
- PDF summary report
- Charts: sentiment, polarity, keyword clouds

8. Conclusion

This project demonstrates the synergy of low-code tools like Octoparse and Python analytics to dissect media

Business Day News Article Analysis Report

Category: Economy | Source: BusinessDay.ng

tone, track editorial focus, and measure shifts in economic narrative from a major Nigerian news outlet.