

BusinessDay Nigeria Economy Article Analysis

1. Project Overview

This project explores news coverage from the Economy category on BusinessDay Nigeria. The goal was to scrape, clean, analyze, and visualize sentiment and keyword trends in economic reporting using a mix of **no-code tools** like Octoparse and **Python-based NLP techniques**.

2. Tools & Technologies Used

Category	Tools/Libraries	Description
Web Scraping	Octoparse, requests, BeautifulSoup	Extracted URLs and article content
Data Handling	pandas, datetime, os	Structured and cleaned scraped data
Text Preprocessing	nltk, re, string	Tokenization, stopword removal, lemmatization
Sentiment Analysis	nltk.sentiment.vader	Classified tone as Positive, Neutral, or Negative
Keyword Extraction	rake_nltk	Extracted high-ranking topic phrases
Visualization	matplotlib, seaborn, wordcloud	Built charts and visual summaries

3. Methodology

✔ Step-by-Step Workflow

1. **Scraping**
 - Collected over 5000 article URLs from BusinessDay’s Economy section using Octoparse.
2. **Article Extraction**
 - Parsed metadata: Title, Author, Date, and Full Content via BeautifulSoup.
3. **Sentiment Analysis**

- Used VADER to classify sentiment and assign polarity scores (-1 to +1).

4. Keyword Summarization

- Applied RAKE on each article to extract key terms and topics.

5. Text Cleaning

- Lowercased text, removed punctuation, lemmatized words, and filtered stopwords.

6. Cleaned Sentiment Analysis

- Re-ran sentiment scoring on cleaned content to assess preprocessing impact.

4. Challenges & Solutions

Challenge	Description	Solution
403 Forbidden errors	Automated scraping was blocked	Used rotating headers & user-agents
Missing/malformed HTML	Some articles lacked authors or markup	Used .get() and try/except safely
Sentiment loss after cleaning	Preprocessing reduced expressiveness	Preserved both raw and cleaned polarity
Noisy keyword results	Irrelevant terms in RAKE output	Extracted keywords from cleaned text only
Inconsistent publication dates	Format varied across articles	Handled using <code>pd.to_datetime(errors="coerce")</code>

5. Visual Insights

◆ Sentiment Distribution

- Majority of articles were **Positive** in tone.
- **Negative** and **Neutral** articles were far fewer.

◆ Monthly Polarity Trends

- Average sentiment polarity remained **consistently high (~0.95+)**.

- Occasional dips aligned with key policy/economic tensions.

◆ **Keyword Analysis**

- Top recurring terms: **CBN, FX, Inflation, Budget, Investment**
- Authors often centered content around consistent economic themes.

6. Key Insights & Recommendations (Elaborated)

Insight	Observation	Recommendation
Positivity Bias	The majority of articles in the Economy section carried a positive sentiment , as reflected by high polarity scores (often > 0.9). This suggests a media tendency to frame economic developments optimistically, even when the underlying data may be neutral or mixed. This could be intentional to foster confidence among investors, policymakers, and the general public.	Leverage this optimism strategically. Organizations and government bodies can align public announcements or policy rollouts with periods of high positivity in the media to boost public reception. However, media outlets should balance optimism with constructive critical analysis to maintain credibility and ensure informed readership.
Recurring Themes	Key economic terms appeared consistently across articles, particularly CBN (Central Bank of Nigeria), Inflation, Foreign Exchange (FX), Budget, and Investment . These recurring keywords indicate the dominant narrative themes in the economy section, showing editorial focus and audience interest in macroeconomic topics.	Capitalize on trending economic themes by developing in-depth series, explainers, or data stories around these topics. This will help increase engagement from a target audience interested in finance, policy, and economic reform. Use SEO strategies to improve visibility and organize content around these high-traffic terms.
Editorial Focus	The analysis showed that certain authors contributed frequently , with recurring bylines and topic preferences. This suggests that some journalists are subject-matter experts	Newsrooms should invest in these recurring contributors by providing access to specialized data, training, or tools. Editors can also track each author’s sentiment patterns,

Insight	Observation	Recommendation
	or have editorial mandates to cover specific economic beats.	engagement metrics, and keyword focus to optimize performance and content diversity . Offering analytics dashboards may encourage more consistent quality and author accountability.
Preprocessing Effect on Sentiment	When articles were cleaned for analysis (removing punctuation, lowercasing, lemmatizing), the overall sentiment polarity slightly decreased . This shows that text preprocessing can dilute emotional or subjective signals that are valuable in sentiment analysis, such as exclamations, negation words, or intensifiers.	It's important to retain both raw and cleaned text versions in sentiment pipelines. For high-stakes analysis (e.g., reputation monitoring or political reporting), consider using context-aware models like BERT, RoBERTa, or LLMs that understand semantic nuance better than rule-based systems like VADER. This ensures more reliable sentiment interpretation and better insight extraction.

7. Deliverables

- updated_scraped.csv – Full dataset with sentiment, polarity, keywords
- SCRAPED.ipynb – End-to-end data pipeline in Jupyter Notebook
- BusinessDay_Economy_Report.pdf – Final write-up of all findings
- BusinessDay_Economy_Presentation.pptx – Slide presentation
- octoparse.img – Screenshot of the Octoparse scraping workflow
- requirements.txt – Python dependencies used

8. Conclusion

This project combines **low-code scraping (Octoparse)** with **Python NLP workflows** to uncover editorial and economic trends in Nigerian media.

It's a **scalable framework** that can be adapted for:

- Real-time news monitoring
- Public policy analysis
- Media bias reporting
- Content strategy & editorial planning