Comprehensive Written Report

Sentiment Analysis of News Articles on Nigeria's Economy (2021–Present)

1. Introduction (Research)

Objective:
 The main goal of this project is to perform **sentiment analysis** on Nigerian economic news articles from **2021 to 2025** to understand the emotional tones expressed in media coverage over time.

The project also seeks to **visualize trends** and **identify dominant narratives** (positive, neutral, negative) while observing **changes year-by-year**.

2. Methodology

2.1 Data Collection and Extraction

- Articles were **scraped directly** from **BusinessDay Nigeria's "BDLead" Section** using requests and beautifulsoup4.

- The scraping process extracted the following key fields from the website:

  - **Title**: The headline of each article.

  - **Author**: The name of the journalist or reporter (where available).

  - **Content**: The full article text.

  - **Date**: The date the article was published.

Scraping Strategy:

- Scraped multiple pages iteratively to cover all articles between **2021 to the present** (2025).

- Implemented respectful scraping practices such as adding delays between requests to avoid server overload.

Data Extraction Process:

- Parsed HTML content to locate articles, titles, authors, publication dates, and body text.

- Cleaned extracted data by removing unwanted HTML tags, whitespace, and formatting inconsistencies.

Post-Scraping Processing:

- Used **intelligent fuzzy matching** (rapidfuzz) during quality assurance to cross-validate titles and ensure no duplicates or missed records.

- Ensured that every extracted article had **all key fields** (Title, Author, Content, Date) before proceeding to analysis.

Final Dataset:

- **~11,000** properly dated and cleaned articles.

- Stored in a structured format (CSV) for subsequent text cleaning, sentiment analysis, and visualization stages.

Step	Action
Data Source	Directly scraped from https://businessday.ng/tag/bdlead/?amp
Tools Used	requests, beautifulsoup4, pandas, rapidfuzz
Fields Extracted	Title, Author, Content, Date

Quality Assurance	Title-content fuzzy matching and missing data checks
Final Dataset Size	~11,000 full articles with complete metadata

2.4 Feature Engineering

- **Year**: Extracted from the Date column for year-by-year analysis.

- **YearMonth**: For monthly tracking of sentiment changes.

2.5 Visualization and Analysis

We performed multiple important visualizations to deeply understand the data:

Visualization	Description
Sentiment Distribution	Bar chart showing total count of Positive, Neutral, and Negative articles overall.
Polarity Score Distribution	Histogram showing how sentiment scores were distributed between -1.0 and +1.0.
Sentiment Over Time (Monthly)	Line chart showing average polarity scores over time by month.
Sentiment Per Year (Average Sentiment Score)	Line chart showing how positivity grew slightly year by year (2021–2025).
Sentiment Category Per Year	Grouped bar chart comparing counts of Positive, Neutral, and Negative articles year by year.
Top 5 Most Positive and Most Negative Articles	Bar charts showing articles with highest positive and negative polarity scores.
Author Sentiment Word Cloud	Word cloud showing authors colored by their average article sentiment.

Word Clouds by Sentiment	Separate WordClouds for Positive, Neutral, and Negative articles showing most common words used in each tone.

3. Key Findings

Sentiment Distribution Findings:

- Majority of articles were **Positive**.
- A significant number were **Negative** .
- A smaller but meaningful portion were **Neutral** (neither strongly positive nor negative).

Year-by-Year Sentiment Analysis:

- **2021–2023**: Sentiment was **stable** with slightly positive average (~0.06).

- **2024**: A **small increase** in average positivity (~0.07).

- **2025**: A **noticeable jump** to about ~0.085 average polarity — the most optimistic year.

Polarity Score Distribution:

- Most articles had polarity scores **between -0.1 and +0.2**.

- Very few articles were extremely negative or extremely positive — meaning most reporting was **moderate and balanced** with a tilt toward optimism.

Top Articles:

- Positive articles discussed **growth, reforms, investments, and recovery**.

- Negative articles highlighted **inflation, insecurity, and economic instability**.

- Neutral articles mostly reported **announcements, official reports, and data without strong emotional tone**.

WordCloud Insights:

- Positive articles showed keywords like: **growth**, **investment**, **development**, **recovery**.

- Negative articles showed keywords like: **inflation**, **unemployment**, **debt**, **corruption**.
  Sentiment by Author:
- Most authors maintained a **neutral to mildly positive** tone.

- No extreme bias toward positivity or negativity observed.

4. Conclusion

Through comprehensive sentiment analysis, it is clear that **Nigerian economic media coverage** has shown a **gradual increase in optimism** between 2021 and 2025.
Although the majority of reporting remained moderately positive, a rising confidence is evident especially in 2024 and 2025.

This trend suggests improving media narratives around economic performance, despite the persistence of challenges like inflation and insecurity.

This analysis can help policymakers, economists, and journalists better understand public sentiment over time.

Full Project Covers:

Stage	Details
Extraction	Titles, Content, Dates, Authors properly merged
Cleaning	Lowercase, punctuation removal, tokenization, lemmatization
Sentiment Analysis	Polarity Scoring, Sentiment Classification
Visualizations	Distribution charts, WordClouds, Top articles, Yearly trends
Reporting	Full findings, interpretation, actionable insights