BUSINESSDAY ARTICLE FINDINGS BY GROUP 5

**BLESSING ALABI**

**RACHEAL WILSON**

**MOYINOLUWA OSO**

**RESEARCH**

Thisresearch is focused on understanding the opinions and emotional tones expressed in Articles related to Nigeria’s Economy from 2021 to present.

The outcome is to provide insights into how the Economic narrative of Nigeria has evolved over time through media coverage.

**METHODOLOGY**

Web scrapping (Extract structured data like Title, Author, Date, content and url)

Data cleaning (Ensures the scraped data is constituent and ready for analysis)

Data storage (Save the scraped data either in csv or excel for further analysis)

Sentiment Analysis (Analyze the sentiment of the article content)

**Tools and libraries used**

1. Selenium used for automating web scraping by interacting with dynamic web pages
2. Undetected-chrome driver: used to by passanti-bot mechanisms like cloudflare challenges.
3. Pandas : used to handle and manipulate the scraped data.
4. JSON: used to save and load checkpoint data for resuming the scraping data.
5. Python standard libraries : os was used to manage file paths and directories, time was used to introduce delays for human like behavior and timestamp checkpoint data, while timestamp was used to generate random delays for scrolling and navigation.
6. Web driver-manager : was used to automatically manage and update the chrome webdriver
7. Exception handling : Ensures the script continues to run even if errors occurs.

**KEY FINDINGS**

The article was scraped from 2021 to present to show the emotional tones expressed in Articles as expressed by the people.

From the results of the analysis its shows that

Sentiment Analysis Results:

Neutral: 5138 articles (63.53%)

Positive: 1997 articles (24.69%)

Negative: 952 articles (11.77%).

Neutral has the highest which is an indication that a large number of people are satisfied with the articles.

**CHALLENGES AND RESOLUTIONS**

**1.**Cloudflare and CAPTCHA : I was blocked severally from accessing the website

**Solution :** I installed undetected chromedriver to bypass cloudflare challenges by making the browser less detectable and also added a manual wait (time.sleep(30))

**2.** Pagination navigation errors : The scrapper failed to load several times due to network issues

**3.**Interrupted excecution when the script was interrupted by closing the terminal it will save the checkpoint and resume from the last processed page when rerun.