OSINT Social Media Analyzer

A comprehensive Python-based OSINT (Open Source Intelligence) tool that analyzes publicly available information from multiple social media platforms and news sources to determine sentiment around specific keywords or topics.

Features

- **Multi-Platform Data Collection**
  - Reddit posts and comments (using PRAW)
  - Google News articles (using BeautifulSoup)
  - RSS feeds from major news sources (using feedparser)
- **Advanced Sentiment Analysis**
  - VADER Sentiment Analysis
  - TextBlob sentiment scoring
  - Positive, Negative, and Neutral classification
  - Compound sentiment scores
- **Rich Visualizations**
  - Pie charts showing sentiment distribution
  - Bar charts for sentiment counts
  - Score distribution histograms
  - Time series analysis
  - Source comparison charts
- **Comprehensive Reporting**
  - HTML reports with visualizations
  - Text-based summary reports
  - JSON data exports
  - CSV data files
- **Web Interface**
  - User-friendly Flask web application
  - Real-time analysis
  - Interactive dashboards
  - Bootstrap-based responsive design

Requirements

- Python 3.8 or higher
- See `requirements.txt` for full list of dependencies

Installation

1. Clone the Repository

```
git clone <repository-url>
cd OSint_P2
```

2. Create Virtual Environment

```
python -m venv venv
```

3. Activate Virtual Environment

Windows:

```
.\venv\Scripts\Activate.ps1
```

Linux/Mac:

```
source venv/bin/activate
```

4. Install Dependencies

```
pip install -r requirements.txt
```

5. Download NLTK Data

```
python -c "import nltk; nltk.download('vader_lexicon'); nltk.download('punkt')"
```

6. Configure Environment Variables

1. Copy `.env.example` to `.env`:

   ```
   copy .env.example .env  # Windows
   cp .env.example .env    # Linux/Mac
   ```

2. Edit `.env` and add your Reddit API credentials:

   - Go to https://www.reddit.com/prefs/apps

- Create a new app (script type)
- Copy the client ID and secret
- Update the `.env` file

Note: Reddit scraping is optional. The tool will work with Google News and RSS feeds even without Reddit credentials.

Usage

Web Application (Recommended)

1. Start the Flask web server:

```
python app.py
```

2. Open your browser and navigate to:

```
http://localhost:5000
```

3. Enter a keyword or hashtag and select data sources
4. Click "Analyze Sentiment" and wait for results
5. View visualizations and download reports

Command Line

Run a quick analysis from the command line:

```
python osint_analyzer.py
```

Or modify the `main()` function in `osint_analyzer.py` to customize your analysis.

Testing Individual Components

Test Reddit Scraper:

```
python scrapers/reddit_scraper.py
```

Test News Scraper:

```
python scrapers/news_scraper.py
```

Test RSS Scraper:

```
python scrapers/rss_scraper.py
```

Test Sentiment Analyzer:

```
python sentiment/analyzer.py
```

Test Visualizations:

```
python visualization/charts.py
```

Libraries Used

1. **praw** - Reddit API wrapper for collecting Reddit posts
2. **BeautifulSoup4** - Web scraping for Google News articles
3. **feedparser** - RSS feed parsing for news sources
4. **vaderSentiment** - VADER sentiment analysis
5. **textblob** - TextBlob sentiment analysis
6. **matplotlib** - Data visualization and charts
7. **seaborn** - Enhanced visualizations
8. **pandas** - Data manipulation and analysis
9. **flask** - Web application framework
10. **nltk** - Natural Language Processing toolkit

Project Structure

```
OSint_P2/
├── scrapers/
│   ├── __init__.py
│   ├── reddit_scraper.py      # Reddit data collection
│   ├── news_scraper.py         # Google News scraping
│   └── rss_scraper.py          # RSS feed parsing
├── sentiment/
│   ├── __init__.py
│   └── analyzer.py             # Sentiment analysis engine
├── visualization/
│   ├── __init__.py
│   └── charts.py               # Chart generation
├── reports/
│   ├── __init__.py
│   └── generator.py            # Report generation
├── templates/
│   ├── index.html              # Home page
```

```
│   ├── results.html           # Results page
│   └── error.html             # Error page
├── static/
│   ├── css/
│   │   └── style.css          # Custom styles
│   └── images/
│       └── charts/            # Generated charts
├── osint_analyzer.py          # Core analyzer engine
├── app.py                     # Flask web application
├── requirements.txt           # Python dependencies
├── .env.example               # Environment variables template
└── README.md                  # This file
```

Academic Implementation Notes

This project demonstrates:

1. **Exception Handling**: Comprehensive error handling for:
    - API rate limits (Reddit)
    - Network connection errors
    - Missing or invalid data
    - Parsing errors
2. **Data Collection**: Three distinct data sources:
    - Reddit API (PRAW)
    - Web scraping (BeautifulSoup)
    - RSS feeds (feedparser)
3. **Sentiment Analysis**: Dual analysis using:
    - VADER (lexicon-based)
    - TextBlob (rule-based)
4. **Visualizations**: Multiple chart types:
    - Pie charts (matplotlib)
    - Bar charts (matplotlib)
    - Histograms (matplotlib)
    - Time series (matplotlib/seaborn)
5. **Reporting**: Comprehensive outputs:
    - Summary statistics
    - Trend analysis
    - Sentiment breakdown
    - Data exports

Important Notes

- **Rate Limits**: Be mindful of API rate limits, especially for Reddit
- **Web Scraping**: Google may block requests if you scrape too aggressively
- **Data Privacy**: This tool only collects publicly available data
- **Academic Use**: Intended for educational and research purposes

Troubleshooting

Reddit Not Working

- Ensure your `.env` file has valid credentials
- Check that your Reddit app is set to "script" type
- The tool will still work with News and RSS if Reddit fails

No Charts Generated

- Check that `static/images/charts/` directory exists
- Ensure matplotlib backend is set correctly
- Check file permissions

Import Errors

- Make sure all dependencies are installed: `pip install -r requirements.txt`
- Download NLTK data: `python -c "import nltk; nltk.download('vader_lexicon')"`

Web App Not Starting

- Check that port 5000 is not in use
- Verify Flask is installed correctly
- Check the console for error messages

License

This project is developed for academic purposes. Please ensure you comply with the terms of service of all platforms you scrape data from.

Acknowledgments

- Adapted from Jose-Sabater/marketeer repository
- Uses VADER Sentiment Analysis (Hutto & Gilbert, 2014)
- Built with Flask, Bootstrap, and modern web technologies

Author

Developed as part of an OSINT analysis project demonstrating social media sentiment analysis techniques.

Support

For issues or questions:

1. Check the troubleshooting section
2. Review the code comments

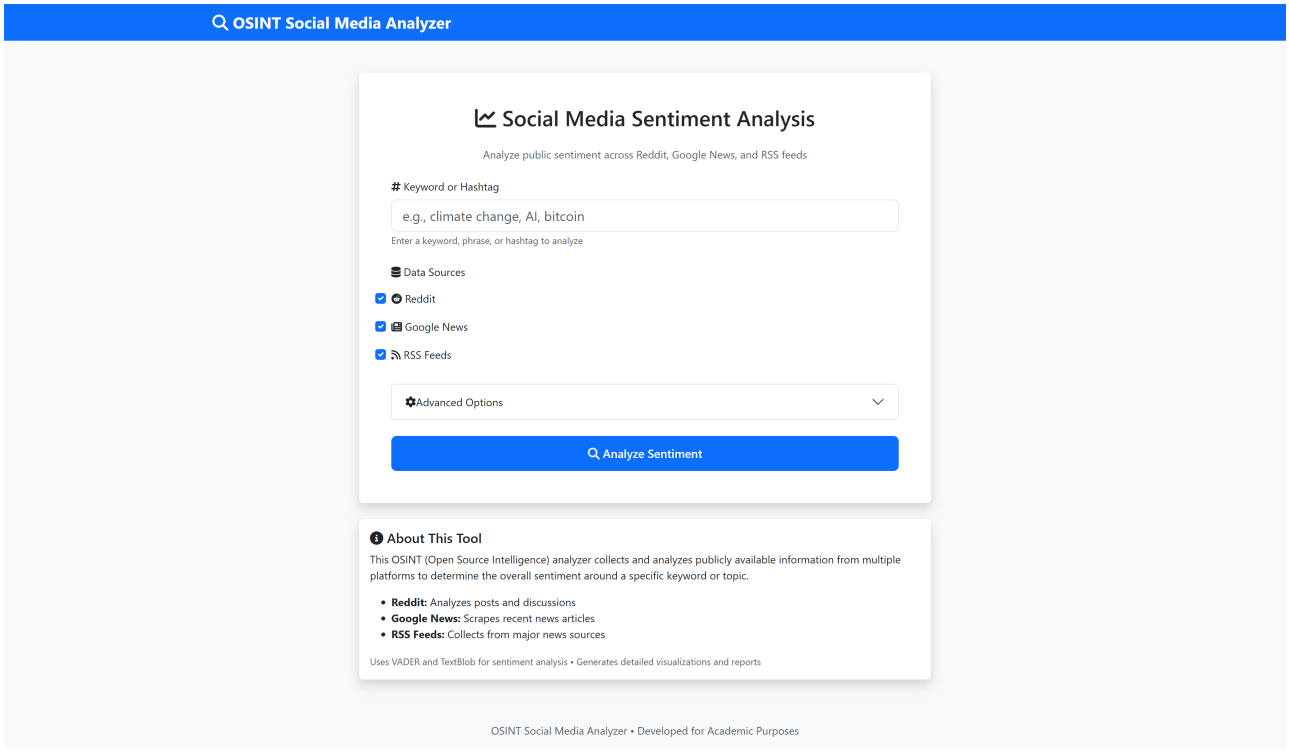
3. Test individual components separately
4. Verify all dependencies are installed

Note: This tool is for educational purposes. Always respect platform ToS and rate limits.

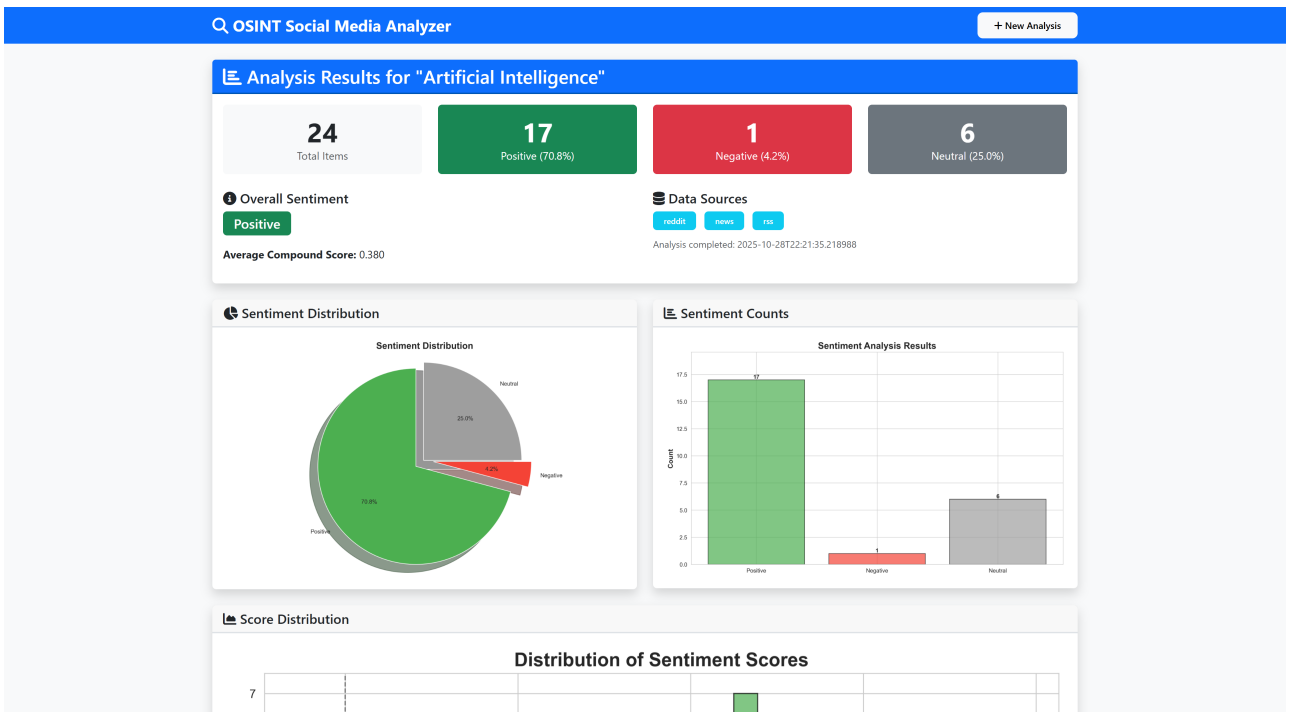
Screenshots

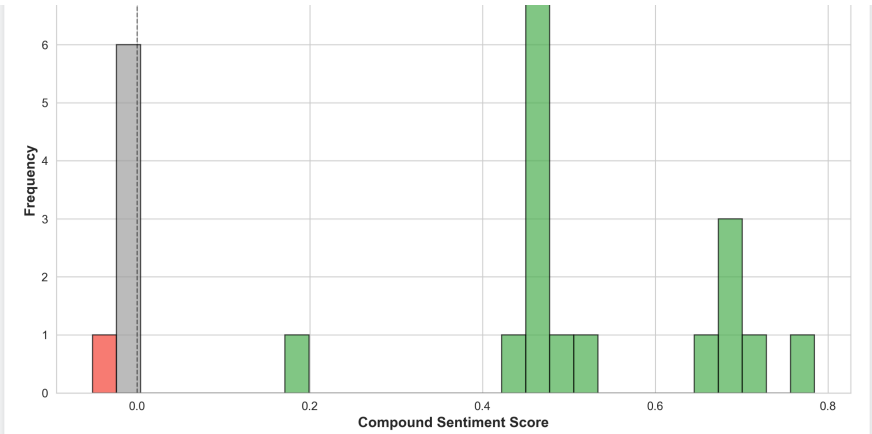
Quick visual examples generated by the project (saved under `Screenshots/`). Browse these while testing or when filing issues to show the current UI/outputs.

- Main web UI

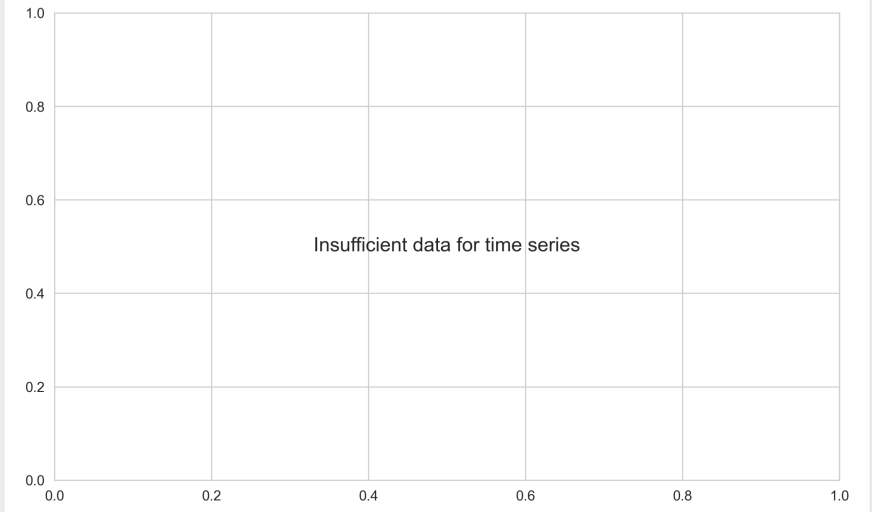


- Example AI analysis report

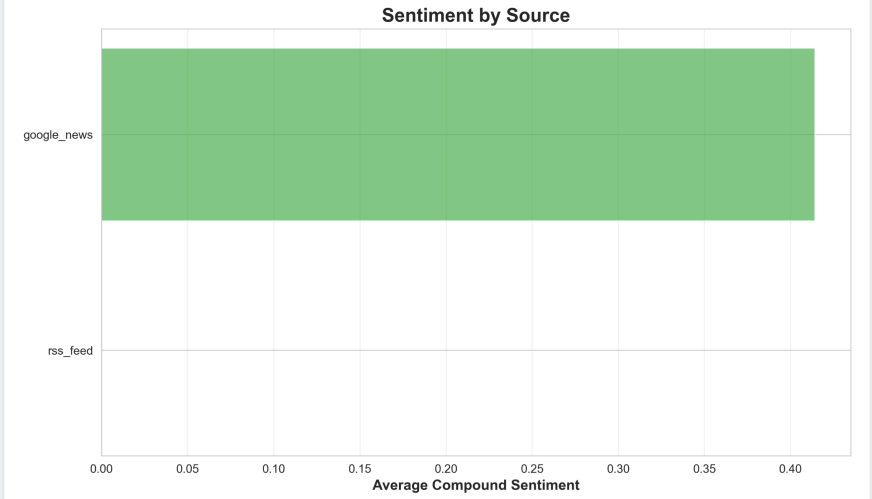




Sentiment Over Time



Sentiment by Source



Detailed Sentiment Breakdown

Metric	Value
Average Positive Score	0.223
Average Neutral Score	0.762
Average Negative Score	0.015
Average Compound Score	**0.380**

+ Analyze Another Keyword | Print Report

OSINT Social Media Analyzer • Developed for Academic Purposes

- Example Bitcoin topic analysis

Analysis Results for "Bitcoin"

22	7	4	11
Total Items	Positive (31.8%)	Negative (18.2%)	Neutral (50.0%)

Overall Sentiment

Neutral

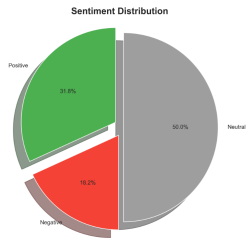
Average Compound Score: -0.025

Data Sources

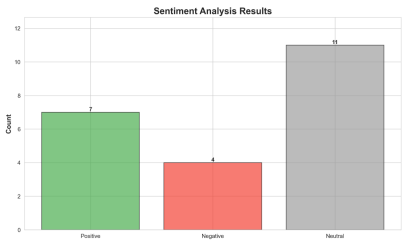
reddit news rss

Analysis completed: 2025-10-28T22:24:26.664969

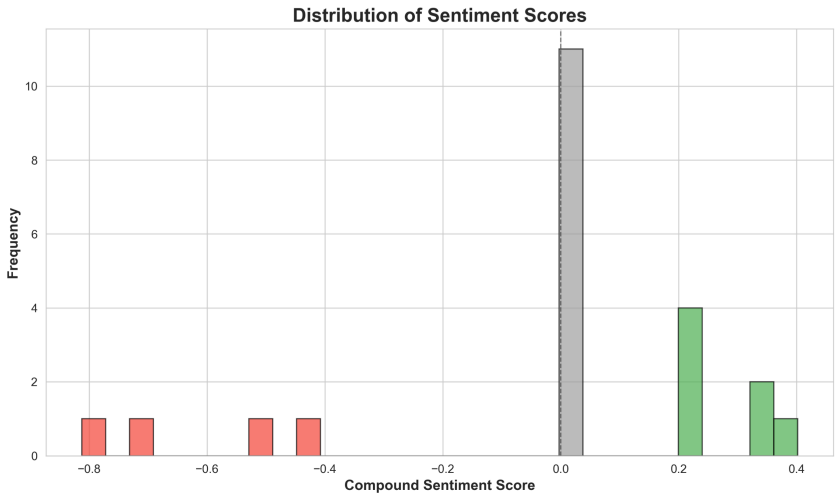
Sentiment Distribution



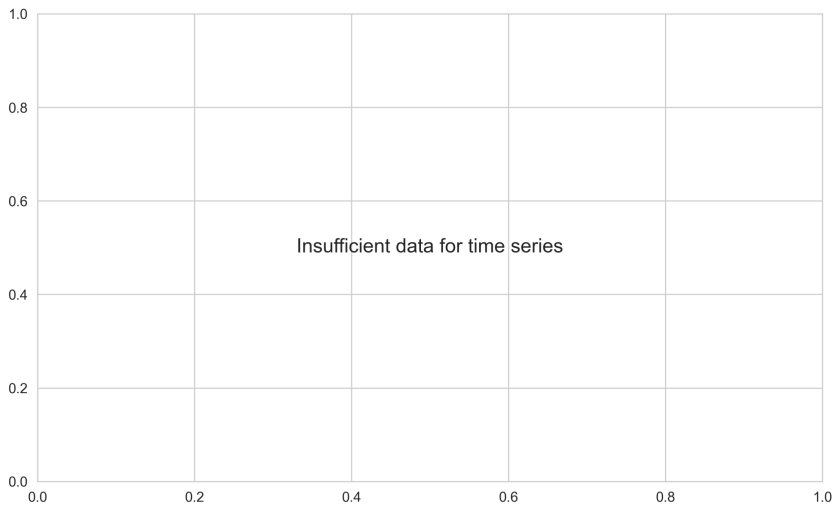
Sentiment Counts



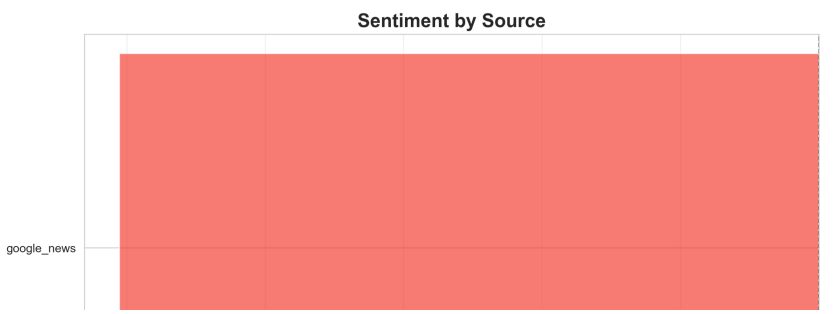
Score Distribution

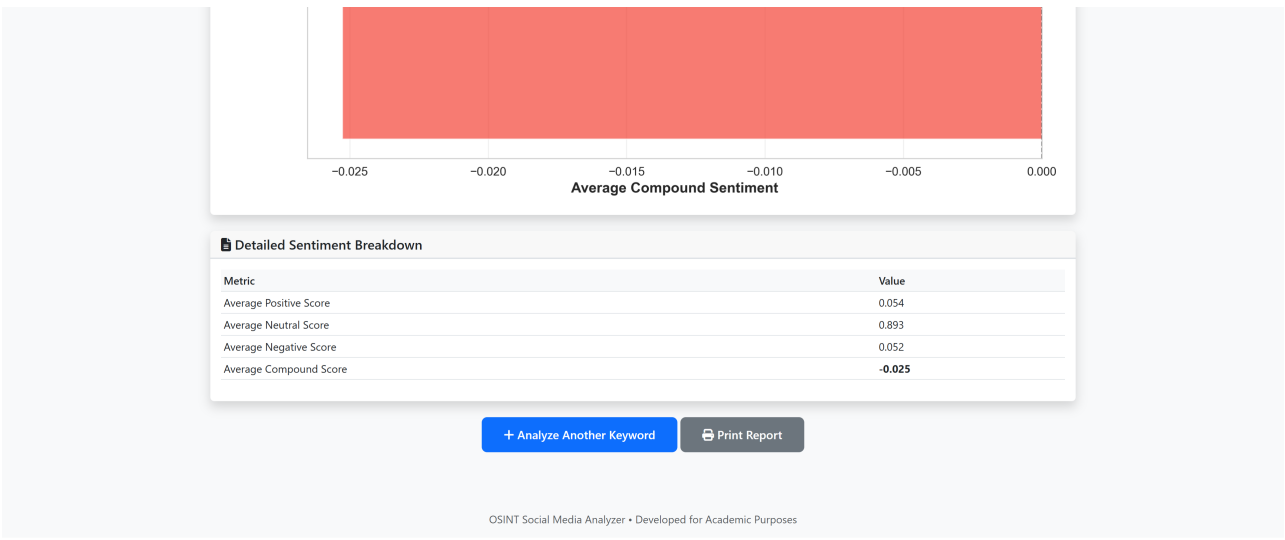


Sentiment Over Time



Sentiment by Source





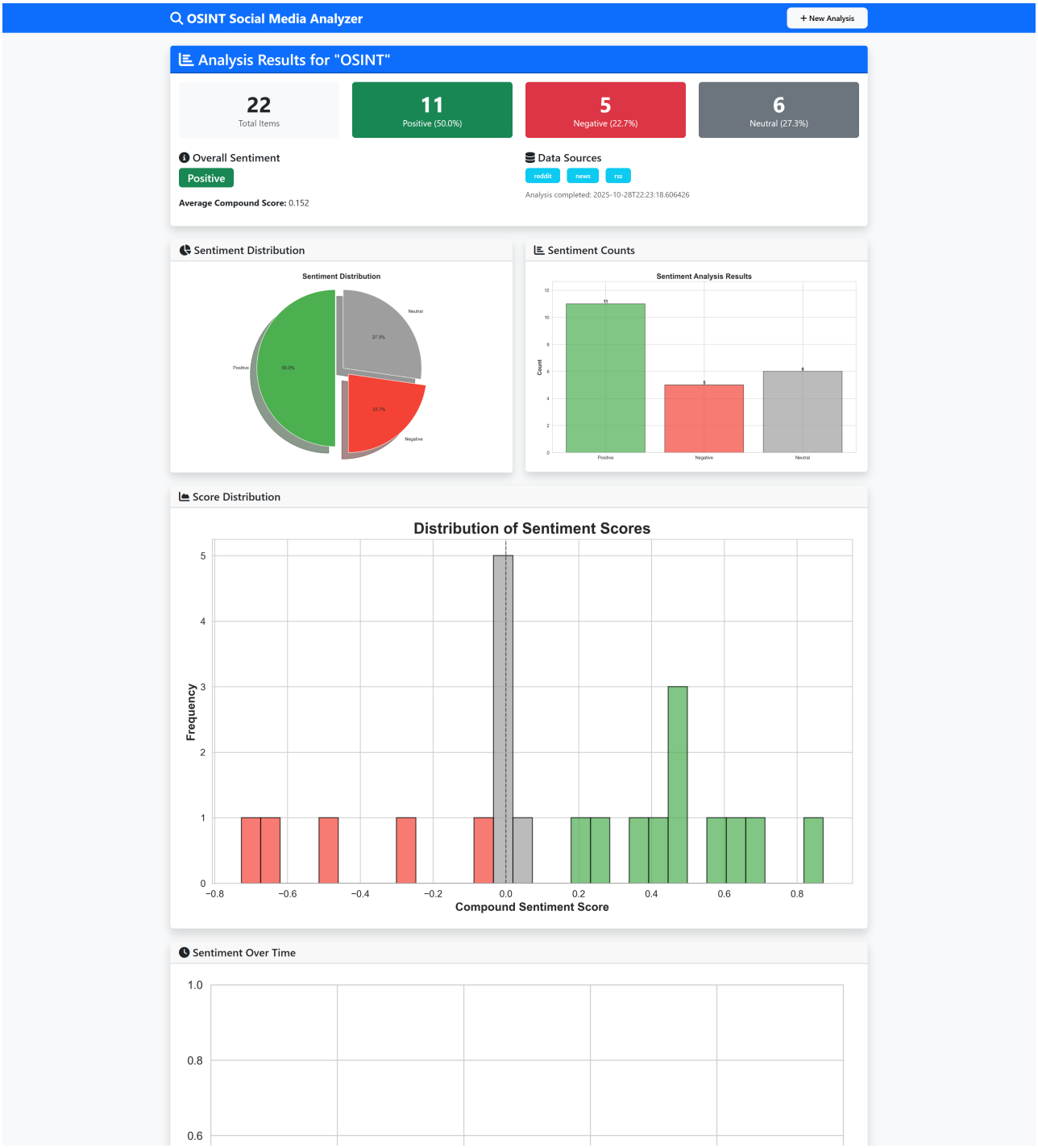
Detailed Sentiment Breakdown

Metric	Value
Average Positive Score	0.054
Average Neutral Score	0.893
Average Negative Score	0.052
Average Compound Score	**-0.025**

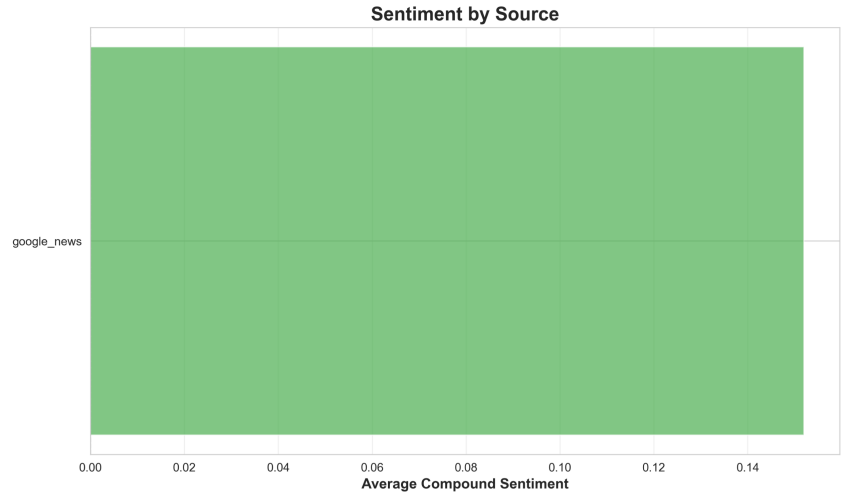
+ Analyze Another Keyword | Print Report

OSINT Social Media Analyzer • Developed for Academic Purposes

- General OSINT run example



Sentiment by Source



Detailed Sentiment Breakdown

Metric	Value
Average Positive Score	0.169
Average Neutral Score	0.744
Average Negative Score	0.087
Average Compound Score	**0.152**

+ Analyze Another Keyword | Print Report