

Project Report : Sentimental and Financial Analysis.

By Lubabah HAMOUCH & Mohammed FAKIR

Module: Business Intelligence

Professor: Dr. KHIATI Wassim

Contents

1. Initial Synopsis.....	5
1.1 Project Overview	5
1.2 Objectives	5
2. Data Collection and Preparation	5
2.1 Data Sources	5
2.2 Data Processing Methodology	5
2.3 Challenges and Solutions.....	6
3. Dashboard Development in Power BI	6
3.1 Dashboard Structure.....	6
3.2 Key Visualizations	6
3.3 Interactive Features.....	7
4. Analysis and Findings.....	8
4.1 Economic Indicator Analysis.....	8
4.2 Stock Performance Analysis	9
4.3 Sentiment Analysis Findings	10
4.4 Correlation Between Sentiment and Market Performance	12
5. Business Implications	12
5.1 Investment Decision Support.....	12
5.2 Risk Management	12
5.3 Strategic Planning	12
6. Technical Implementation	12
6.1 Tools and Technologies Used.....	12
6.2 Data Model.....	13
7. Limitations and Future Improvements.....	13
7.1 Current Limitations	13
7.2 Future Enhancements.....	13
Conclusion	14
References.....	15
Appendices.....	15

Table of Figures

Figure 1: Dashboard on Power BI	6
Figure 2: News Data Analysis on Power BI	7
Figure 3: Stock Market Analysis Page on Power BI	8
Figure 4: Economic Indicator Analysis by Country	8
Figure 5: Trading Volumes Chart on Studied Shares	9
Figure 6: Daily Return of Studied Shares	9
Figure 7: Close Price Evolution by Shares	10
Figure 8: Volatility Index for Studied Shared.....	10
Figure 9: Score Sentiment by News Source	11
Figure 10: Sentiment Analysis Results Comparison by Source	11
Figure 11: Average Twitter Sentiment Score Over Time	11

Introduction

Today's world, being extremely globalized, has its financial markets increasingly controlled by a broad range of global events. The political, economic, or social events have far-reaching implications for stock prices. Understanding the ways these events influence financial markets requires sophisticated tools that can manage large amounts of data and provide quality insights. This project aims to utilize the potential of sentiment analysis, AI, and big data in predicting the impact of international events on stock prices. Through the study of different sources of data such as news reports, social media sites, and economic indicators, the project aims to gain a stronger understanding of how sentiment meets market trends.

The project aims primarily to create a model that is able to forecast stock price movement based on sentiment both from general public opinion and historical data. This would involve combining several sources of data: financial data such as the stock price of major companies and market indices; sentiment data that is extracted from text-based content, i.e., news articles and social media posts, and macroeconomic indicators.

Data collection was extensive and taken from many sources. Financial data were gathered from established market indices such as the S&P 500, Dow Jones, and specific technology stocks. Sentiment data were gathered from leading news outlets such as Bloomberg, Reuters, and others, as well as from social media platforms such as Twitter and Reddit, where public reaction and opinion concerning world events are likely more instant and polarized. Additionally, economic statistics such as inflation rates, employment figures, and GDP growth were incorporated to offer some background to the examination. The data sources were then manipulated and merged such that all data points were synchronized with regard to time and pertinence.

Data preparation involves several important steps. Raw data were initially pre-cleaned to remove duplicates, correct missing values, and standardize formats for consistency across the different datasets. Sentiment analysis utilized Natural Language Processing NLP techniques that classified the text data into three sentiment groups: positive, negative, and neutral. These sentiment scores were subsequently integrated with economic and financial data to provide a detailed analysis of the correlation between sentiment trends and the evolution of stock price over time. As a result of the properties of the financial lexicon and the need for greater accuracy, domain-specific dictionaries were employed to make fine adjustments to the sentiment classification such that the analysis provided a more accurate reflection of the nuance of financial language.

In the next sections, we will cover the technicalities of the data collection procedure, sentiment analysis, and dashboard development, then proceed to analyze the insights derived from the data. From this project, our goal is to demonstrate the advantage that can be gained from integrating traditional financial analysis with modern sentiment analysis techniques to develop a better-rounded framework for understanding and predicting market behavior.

1. Initial Synopsis

1.1 Project Overview

Global events and mass opinion heavily influence financial markets. Our project tries to quantify these forces by collecting data from varied sources, sentiment analysis, and plotting correlations among events, sentiment, and stock price movement.

1.2 Objectives

- Gather and examine financial data as well as sentiment data from news and social media; use natural language processing methods to gauge textual sentiment; and explain the relationships between sentiment indicators and shifts in stock prices.
- Develop interactive visualizations that easily illustrate these relationships and offer useful information for financial analysis and investment decision-making.

2. Data Collection and Preparation

2.1 Data Sources

Our project incorporated data from various sources:

- Financial Market Data: Historical stock prices of big-cap tech stocks (AAPL, GOOGL, MSFT) and market index (SPY)
- Economic Indicators: Inflation and unemployment rates of major economies (China, India, United Kingdom, United States)
- News Articles: Content from influential economic news websites like BBC News, Business Insider, Bloomberg, and others
- Social Media: Sentiment data derived from Twitter and Reddit comments related to financial markets

2.2 Data Processing Methodology

The pipeline for data processing included the following:

Web Scraping: Automated collection of social media updates and news stories

Data Cleaning: Removing duplication, filling in missing values, and format standardization

Sentiment Analysis: Applying NLP techniques to identify if the text is positive, negative, or neutral

Data Integration: Merging of financial data with sentiment scores from timestamps

Feature Engineering: Derivative features such as averages of sentiment scores and volatility measures are developed

2.3 Challenges and Solutions

During the data preparation phase, we encountered various challenges:

- Dealing with huge volumes of social media data required optimization of our extraction algorithms
- Enhancing sentiment analysis accuracy for financial terms using domain-specific lexicons
- Appropriate alignment of news events and market reactions with respect to market hours

3. Dashboard Development in Power BI

3.1 Dashboard Structure

Our Power BI solution consists of four main dashboard pages:

1. Analysis of Economic Indicators: Comparison of national economic performance
2. Stock Market Performance: Stock price movement and volume analysis
3. Sentiment Analysis: Visualization of sentiment trends on different news channels and social media
4. News Sentiment Distribution: Detailed news sentiment distribution by source and time

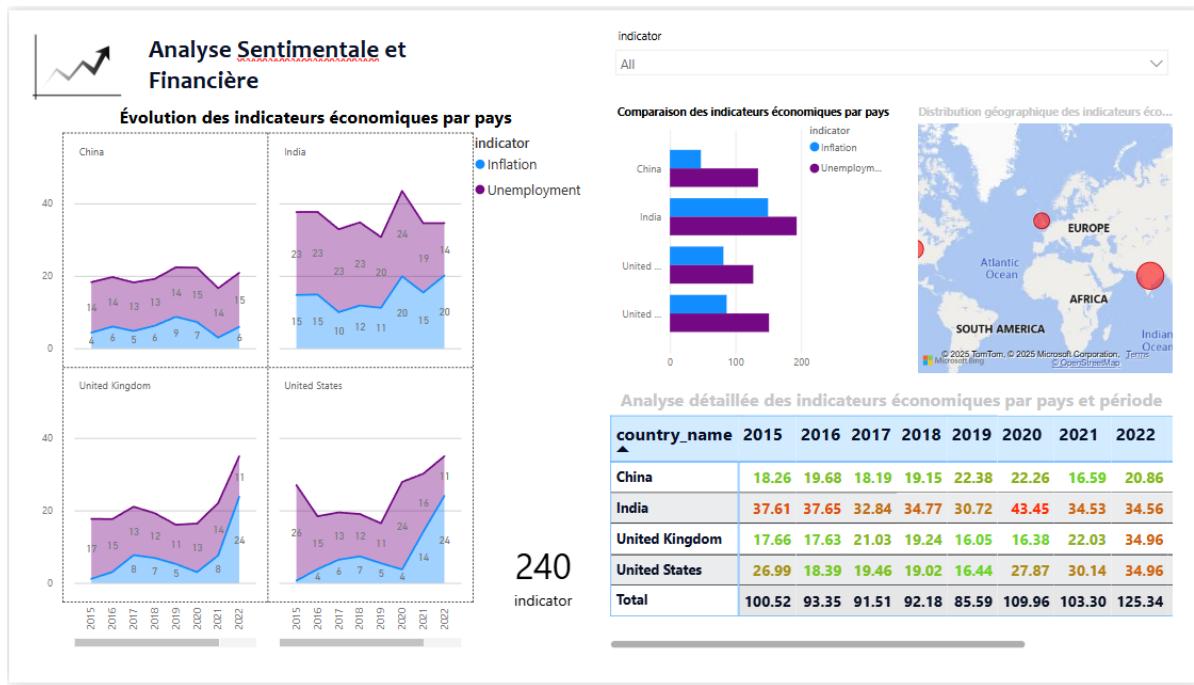


Figure 1: Dashboard on Power BI

3.2 Key Visualizations

The dashboard integrates different forms of visualization:

- Area Charts: Showing the evolution of economic indicators by time per nation
- Bar Charts: Comparison of sentiment scores by different sources
- Line Charts: Trending stock price performance and sentiment scores over time
- Donut Charts: Displaying the distribution of sentiment categories in news articles
- Maps: Displaying the geographical distribution of economic indicators
- Tables: Showing detailed numerical data for precise scrutiny

3.3 Interactive Features

Our Power BI solution includes some interactive elements:

- Time Sliders: Allowing users to look at specific date ranges
- Filters: Allowing filtering based on source, sentiment category, country, and stock ticker
- Drill-down Capabilities: Providing full views by selecting specific data points
- Cross-filtering: Allowing selection in one visualization to filter other visualizations for contextual analysis

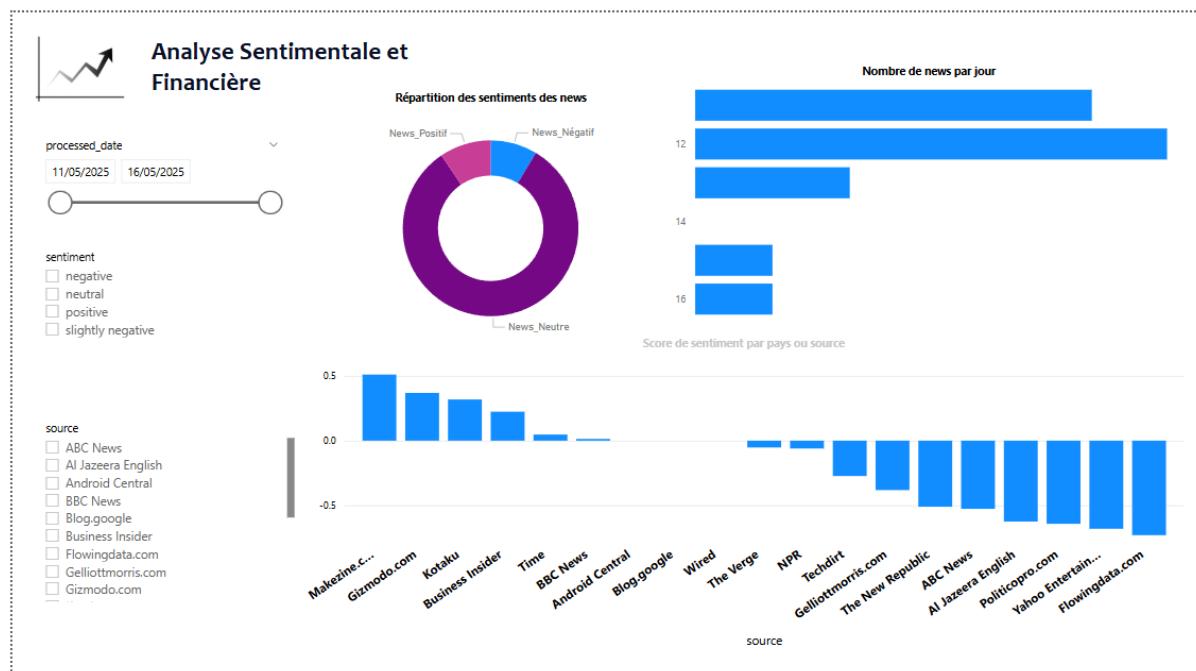


Figure 2: News Data Analysis on Power BI

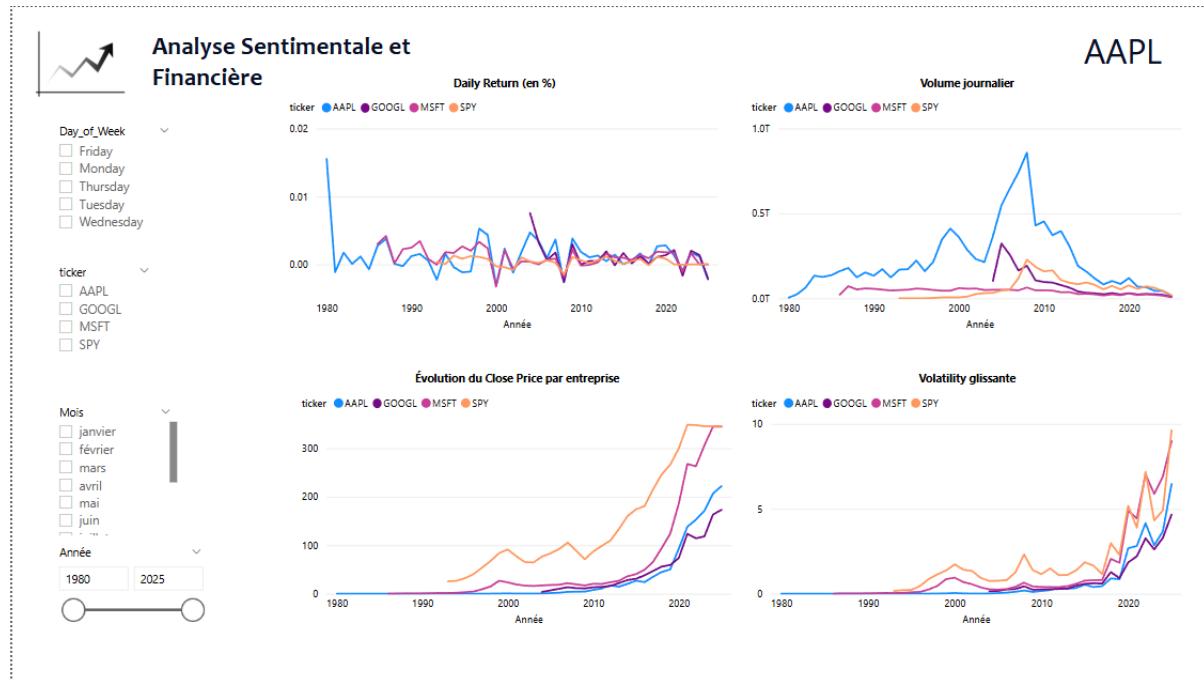


Figure 3: Stock Market Analysis Page on Power BI

4. Analysis and Findings

4.1 Economic Indicator Analysis



Figure 4: Economic Indicator Analysis by Country

- China and India have contrasting inflation and employment trends compared to Western countries
- The year 2020 was a turning point for economic indicators in all the countries under study
- The US and UK had the same trend in the pattern of mildly decreasing unemployment over the considered period

4.2 Stock Performance Analysis

- AAPL captured the highest volatility in trading volumes among studied shares
- Technology shares captured consistent growth in closing prices since 2010 with higher acceleration after 2020
- Volatility of stocks increased significantly during times of economic uncertainty

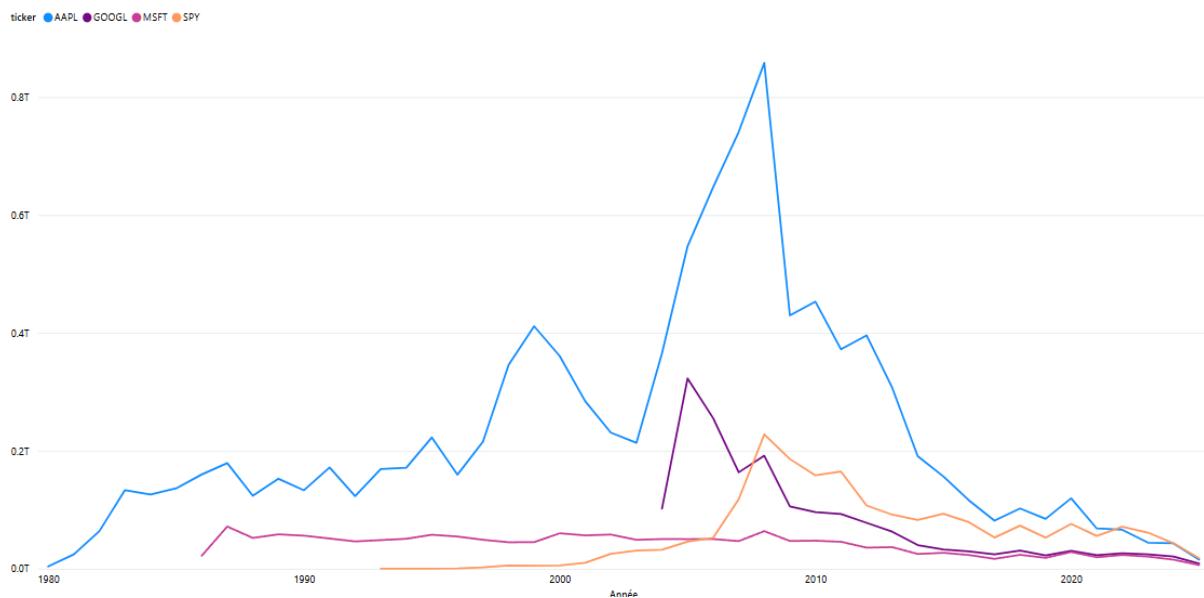


Figure 5: Trading Volumes Chart on Studied Shares

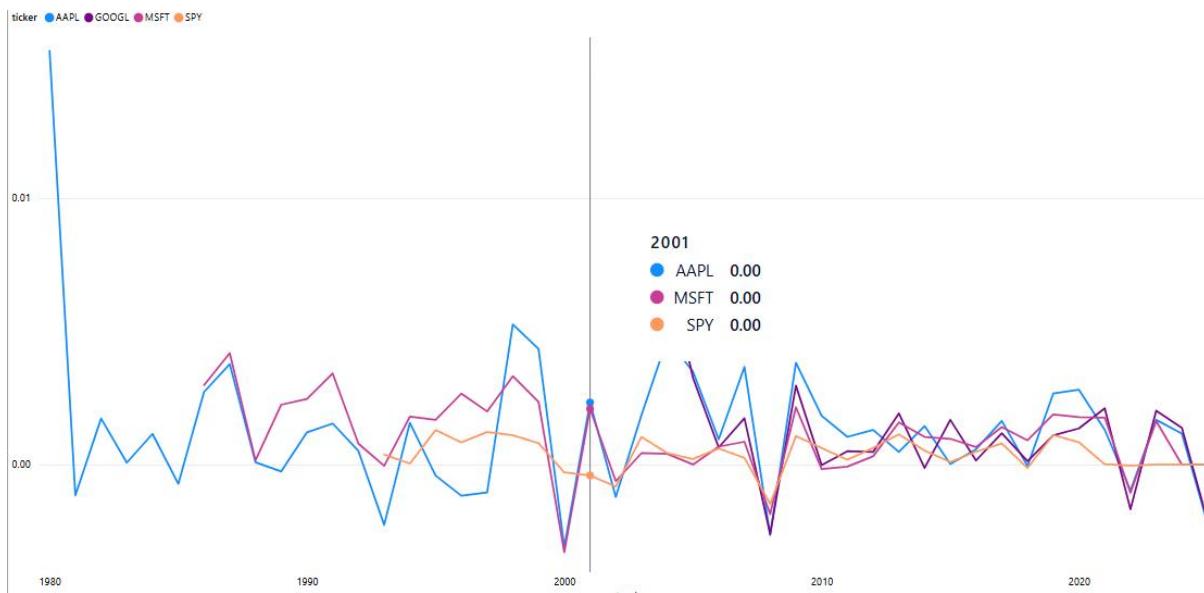


Figure 6: Daily Return of Studied Shares

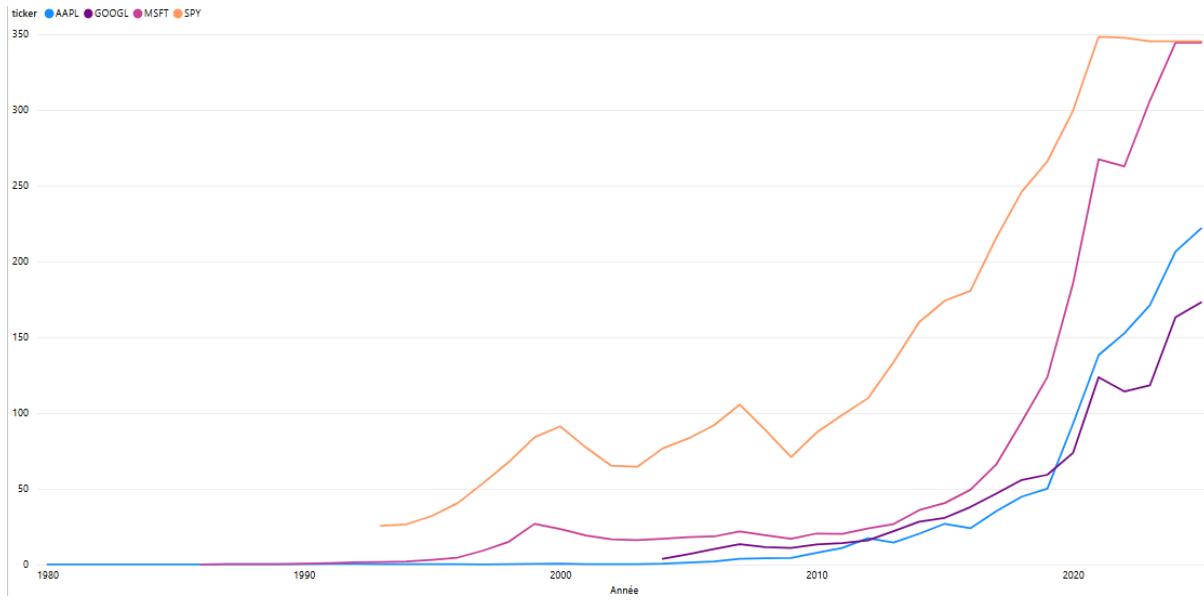


Figure 7: Close Price Evolution by Shares

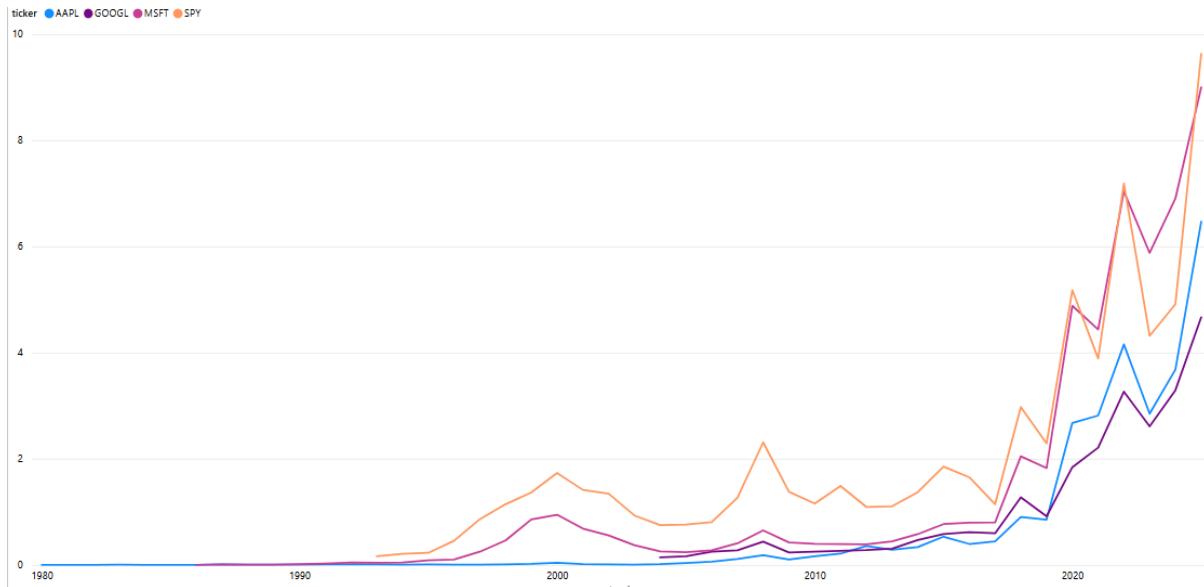


Figure 8: Volatility Index for Studied Shared

4.3 Sentiment Analysis Findings

- News media generally evidenced a predominantly neutral sentiment in their reporting
- Social media platforms recorded more polarized sentiment than mainstream news sites
- Reddit recorded a higher rate of neutral sentiment (0.73 average) among the sites
- Twitter sentiment recorded an overall positive trend during the May 11-16 analysis period

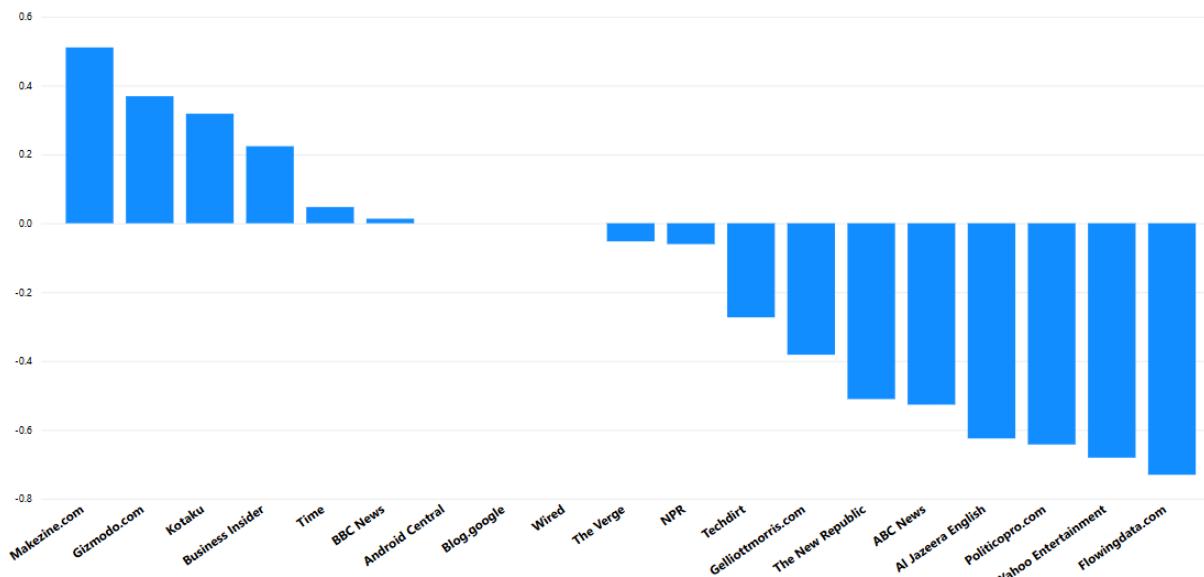


Figure 9: Score Sentiment by News Source

● Average of Positif ● Average of Négatif ● Average of Neutre

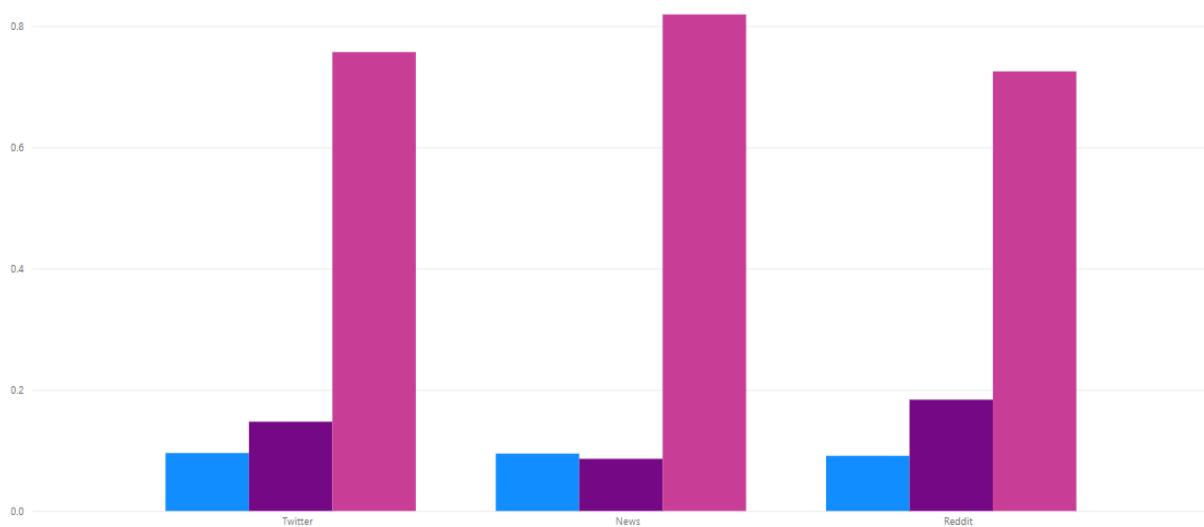


Figure 10: Sentiment Analysis Results Comparison by Source

Score de sentiment moyen dans le temps Twitter



Figure 11: Average Twitter Sentiment Score Over Time

4.4 Correlation Between Sentiment and Market Performance

- Short-term market volatility was usually preceded by highs in negative news sentiment
- Social media showed consistent positive sentiment patterns parallel to stable market gains
- Alternative news sources indicated varying levels of sentiment correlation with market directions

5. Business Implications

5.1 Investment Decision Support

The dashboard provides substantial investment decision support by:

- Recognizing potential market reaction to forming news sentiment trends
- Highlighting economic indicators and individual stock performance relationships
- Providing early warning signals through sentiment analysis before the market reacts

5.2 Risk Management

The sentiment analysis module offers risk management benefits by:

- Gauging market sentiment during periods of high volatility
- Providing source-specific sentiment analysis to determine information reliability
- Following sentiment evolution to anticipate potential market corrections

5.3 Strategic Planning

Organizations can utilize our findings for strategic planning by:

- Understanding how economic indicators influence their own sector
- Monitoring trends in sentiment of their own industry
- Predicting market reactions to major company announcements

6. Technical Implementation

6.1 Tools and Technologies Used

- Data Collection: Python scripts and BeautifulSoup and Selenium for web scraping
- Data Processing: Python and Pandas for data cleaning and manipulation
- Sentiment Analysis: NLTK and TextBlob libraries for NLP-based sentiment analysis
- Data Visualization: Microsoft Power BI for the dashboard

- Data Storage: Both CSV file and SQL database for structured data storage

6.2 Data Model

The data model behind relates:

- Time-based dimensions to analyze by time
- Geographical dimensions to compare by region
- Sentiment score measures to news sources
- Financial metrics to stock tickers and time periods

7. Limitations and Future Improvements

7.1 Current Limitations

- Sarcasm and contextual nuance are not considered in sentiment analysis
- Historical data is limited to narrow date ranges and stock baskets
- The model does not incorporate all economic indicators

7.2 Future Enhancements

- Levelling of advanced machine learning models in order to more accurately predict sentiment
- Wider data sources that cover more international markets and currencies
- Development of predictive analytics to forecast market movements on the basis of sentiment trends
- Integration with real-time data feeds for instant sentiment and market analysis

Conclusion

In brief, this Business Intelligence project has untangled the dynamics of global events, public opinion, and financial market volatility and converted raw data into actionable intelligence. Our Power BI dashboard is more than a collection of visualizations—an interactive dashboard allows users to discover how sentiment, either positive or negative, has a direct impact on stock prices. By combining sentiment analysis and financial data, we've created an easier method of monitoring the effect of global events on markets, with a clearer, more interactive means of forecasting trends.

As it further matures and incorporates additional sources of information, this tool can be a foundation of financial analysis, enabling analysts and investors to unlock the mysteries of the markets with greater confidence. What was a confusing problem before is now an elegant, data-based solution—making the wild world of the financial markets somewhat more comprehensible.

References

- Financial data sources: Yahoo Finance.
- Economic indicators: World Bank, IMF
- Sentiment analysis techniques: Natural Language Processing with Python (TextBlob, Vadersentiment...)

Appendices

Appendix A: Data Dictionary

- **Economic Indicators Dataset:**
 - country_name: Country identifier
 - year: Year of measurement
 - inflation: Annual inflation rate (%)
 - unemployment: Unemployment rate (%)
 - GDP
- **Stock Data Dataset:**
 - ticker: Stock identifier (AAPL, GOOGL, MSFT, SPY)
 - date: Trading date
 - open: Opening price
 - close: Closing price
 - volume: Trading volume
 - return: Daily return (%)
- **Sentiment Dataset:**
 - source: News or social media source
 - processed_date: Date of content publication
 - sentiment: Classified sentiment (positive, negative, neutral, slightly negative)
 - sentiment_score: Numerical sentiment score (-1 to 1)