

# Interim Report on Task 1: Exploratory Data Analysis and Sentiment Analysis of Financial News

## Introduction

In this interim report, I summarize my progress on Task 1 of the Financial Analytics Challenge, which involved conducting exploratory data analysis (EDA) and sentiment analysis on a dataset of financial news articles. The goal was to analyze publication trends, assess sentiment in headlines, and evaluate the contributions of various publishers to the news landscape.

## Methodology

### 1. Data Loading and Preparation:

- I loaded the dataset from a CSV file containing key columns such as `headline`, `date`, and `publisher`.
- The `date` column was converted to a datetime format to facilitate time series analysis. I ensured that any parsing errors were handled by using `errors='coerce'`, which helped identify problematic entries.

### 2. Exploratory Data Analysis (EDA):

- **Publication Trends:** I resampled the data to count the number of articles published daily. This analysis revealed fluctuations in publication frequency, indicating periods of heightened news activity.
- **Publisher Analysis:** I counted articles by publisher to determine which ones contributed most frequently to the dataset, highlighting key players in financial news reporting.

### 3. Sentiment Analysis:

- I implemented sentiment analysis using the TextBlob library, which provided polarity scores for each headline. This allowed me to quantify the sentiment expressed in the articles.
- The distribution of sentiment scores was visualized using histograms, offering insights into the overall tone of financial news.

## Initial Findings

- The analysis of publication trends indicated that certain dates experienced spikes in article counts, potentially correlating with significant market events or economic announcements.
- The publisher analysis showed that a small number of publishers dominate the news landscape, suggesting their reporting may significantly influence market sentiment.

- The sentiment analysis revealed a balanced distribution of sentiment scores across headlines, with a slight inclination towards positive sentiment.

## Challenges Encountered

- **Performance Concerns:** Initial attempts at processing large datasets for sentiment analysis resulted in longer execution times. I recognized the need to explore more efficient libraries or methods for larger datasets in future tasks.

## Conclusion

The progress made in Task 1 has established a solid foundation for further analyses in subsequent tasks. By successfully performing EDA and sentiment analysis, I have gained valuable insights into how financial news is disseminated and its potential impact on market behavior. Moving forward, my focus will shift to quantitative analysis using additional financial data and technical indicators as outlined in Task 2. This report will be updated to reflect further progress as I advance through the project.