# Project Proposal: Real-Time Stock Market Sentiment Analyzer and Price Tracker

## Problem statement formation

* Financial analysts lack a unified, automated system that correlates real-time sentiment with stock price fluctuations.
* Manual tracking of news and prices is inefficient and lacks scalability.
* There is a need to build a reproducible and scalable pipeline that connects these data sources and enriches them with intelligent signals.

## Context

In today’s fast-paced markets, investor sentiment derived from financial news and social chatter can significantly influence stock prices. Institutional investors increasingly rely on real-time analytics to make data-driven decisions. This project demonstrates how modern data engineering can support such needs by enabling real-time ingestion, enrichment, and analytical modeling on the cloud.

## Criteria for success

* Real-time data successfully ingested and persisted
* Sentiment accurately assigned using NLP models
* Strong correlation or insights visible through analytics
* Power BI dashboard displays real-time or near-real-time data
* All components deployed on Azure and version-controlled in GitHub

## Scope of solution space

* End-to-end pipeline with Delta Lake layers (Bronze → Silver → Gold)
* Sentiment-scored news data aligned with stock price trends
* Visual dashboard showing sentiment-impact on price
* Documentation, README, and architecture diagram
* Optional: Deployed ML model for price movement prediction

## Data source(s)

|  |  |  |  |
| --- | --- | --- | --- |
| Category | Source | Update Frequency | Format |
| Stock Prices | Polygon.io or Yahoo Finance | Real-time or minute-level | JSON/CSV |
| Financial News | NewsAPI.org or GNews | Real-time | JSON |

## Proposed architecture for the solution and rationale behind it

A diagram of a software company

AI-generated content may be incorrect.

## Choice of technology for the solution and rationale

* **Ingest Data**: Collect news and stock prices via API and send them to Event Hubs.
* **Stream Processing**: Use Azure Databricks to read from Event Hub and write to Bronze Delta tables.
* **Clean and Enrich**: Transform raw data into Silver tables; apply sentiment analysis.
* **Model**: Correlate news sentiment with price changes and store in Gold tables.
* **Visualize**: Present insights using Power BI and/or Snowflake dashboards.
* **Optional ML**: Predict future price movement based on sentiment trends