

# Project Documentation: Sprint 1 Completion Report

## Real-Time Industry Insight & Strategic Intelligence System

**Phase:** Sprint 1 – Data Sourcing & Handling (Weeks 1-2)

**Team:** Gopichand, Anshika, Janmejay

**Date Completed:** October 30, 2025

### I. Executive Summary

Sprint 1 successfully delivered the **foundational data pipeline** for the entire strategic intelligence system. The core objective was to establish robust, multi-source data collection, cleaning, and structuring.

**Key Outcome:** We developed a functional, interactive notebook (Project\_Sprint1.ipynb) that accepts any company name (e.g., Netflix, Infosys, Rakuten, TCS) and immediately executes the data lifecycle: **Extraction** (from multiple sources), **Transformation** (via pandas), **Analysis** (using a Sentiment POC), and **Visualization**. This effort not only completed Milestone 1 but also delivered a working prototype for the first part of **Module 1 (Market and Sentiment Analysis Engine)**.

### II. Milestone 1 Fulfillment & Status

The tasks defined in Milestone 1 were successfully completed, establishing the necessary infrastructure and data flow.

Official Milestone 1 Task	Deliverable in Project_Sprint1.ipynb	Fulfillment Status
<b>Set up the environment</b> for market data collection (e.g., integration with Twitter API).	All dependencies (pandas, matplotlib, yfinance, wikipedia, etc.) installed. <b>Alpha Vantage API Key</b> acquisition and setup confirmed for reliable stock data in future modules.	<b>Complete</b>
<b>Train team members</b> on using LLMs for sentiment analysis.	Implemented <b>TextBlob</b> as a <b>Proof-of-Concept (POC)</b> to validate the sentiment extraction logic before integrating high-fidelity LLMs (GPT/LLaMA) in Sprint 2.	<b>Complete (Logic Validated)</b>

<b>Begin collecting initial data</b> from news articles, social media platforms, etc.	Implemented dynamic data ingestion from four distinct sources aggregated into structured DataFrames.	<b>Complete</b>
---	--	-----------------

### III. Technical Deep Dive: The Data Pipeline

The project is built on a four-stage ETL (Extract, Transform, Load/Analyze) process, demonstrated for a dynamic user input asset.

#### 1. Data Extraction (E) – Multi-Source Aggregation

The system successfully collects data from the following sources, addressing the need for multi-modal intelligence:

Source	Data Collected	Tool Used	Strategic Rationale
<b>Financial Data</b>	Historical closing prices and volume data.	yfinance (POC) / <b>Alpha Vantage API Key</b> (Future)	Essential for <b>Trend and Positioning Forecasting System (Module 2)</b> .
<b>Contextual Data</b>	Company summary and background.	wikipedia	Provides immediate competitive context for reports.
<b>News/Market Reports</b>	Current headlines and links.	feedparser (Google News RSS)	Real-time event identification for the <b>Opportunity and Threat Alert Module (Module 3)</b> .
<b>Social Media Data</b>	50 <b>Synthetic Tweets</b> generated.	Python scripts	Validates the system's ability to handle unstructured social text for sentiment analysis.

#### 2. Data Transformation (T) – Cleaning and Structuring

- **Tool:** pandas
- **Process:** All extracted raw data is cleaned, normalized, and converted into labeled, time-indexed **pandas DataFrames**. This standardized structure is critical for downstream machine learning models (LLMs and predictive analytics).

### 3. Preliminary Analysis (A) – Sentiment Engine POC

- **Tool:** TextBlob
- **Process:** Applied sentiment analysis to the news and synthetic social data to categorize each entry as **Positive, Negative, or Neutral**.
- **Outcome:** This step validates the core analytical logic of **Module 1**, proving the system can successfully process and label text data based on market mood.

### 4. Insight Visualization (L) – Foundational Dashboard Elements

- **Tool:** matplotlib.pyplot
- **Outcome:** Three key insights charts were generated, forming the baseline for the **Strategic Insight Dashboard (Module 4)**:
  1. **Sentiment Distribution:** Overall market mood analysis.
  2. **Sentiment Trend:** Visualizing the change in sentiment over the dataset.
  3. **Market Price Trend:** Tracking the closing price over time, allowing for visual correlation with sentiment.

## IV. Conclusion and Transition to Sprint 2

**Sprint 1 is officially complete.** The foundation is robust, secure, and dynamic. We have proven the capability to dynamically ingest and analyze data from multiple sources.

### Transition to Sprint 2 (Weeks 3-4):

The focus will now shift from foundational data handling to advanced, **AI-powered intelligence**, fully realizing **Module 1 (Market and Sentiment Analysis Engine)** and beginning **Module 2 (Trend Forecasting)**.

Sprint 2 Target Modules	Key Tasks	Strategic Objective
<b>Module 1 (Analysis Engine)</b>	Replace TextBlob POC with <b>direct LLM integration (GPT/LLaMA)</b> for high-accuracy, nuanced sentiment analysis.	Elevate intelligence quality beyond basic polarity scoring.
<b>Module 2 (Trend Forecasting)</b>	Begin building predictive models using historical price and volume data sourced via the <b>Alpha Vantage API</b> .	Introduce predictive analytics to forecast shifts in competitive positioning.
<b>Module 3 (Alert Module)</b>	Design the structure for monitoring critical sentiment/price changes to trigger alerts.	Lay the groundwork for real-time Slack notifications.