

Stock Price and News Sentiment Analysis

By Anuj, Archita and Dimple (Mandelbulb Technologies)

Project Overview

Our project is a real-time analytics platform that tracks US news events and their impact on stocks. Using advanced NLP algorithms, it analyzes news articles and social media posts to gauge market sentiment, empowering investors to make informed trading decisions.

This project aims to analyze the sentiment of US news related to 10 US stocks and potentially correlate it with the stock prices. It is divided into two phases:

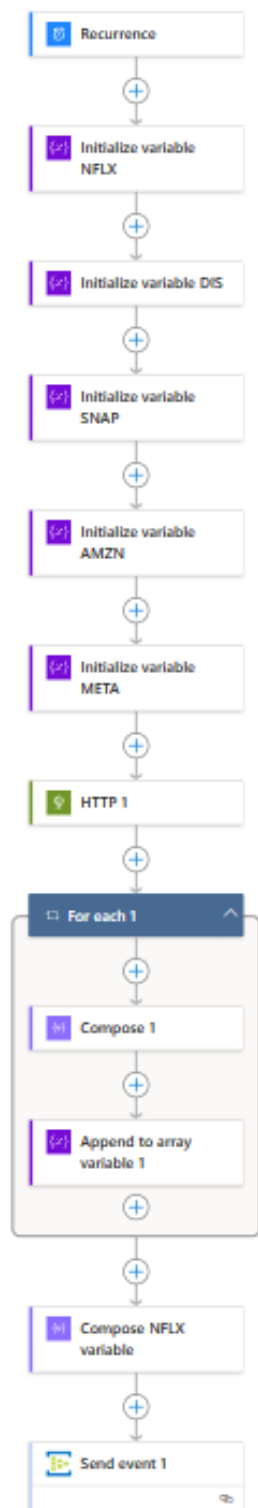
Phase 1: Stock Price Data Ingestion

1.1.Data Source: Twelve Data API (<https://twelvedata.com>)

- Twelve Data offers a programmatic interface (API) for accessing various financial data, including real-time and historical stock prices.
- You'll need to sign up for a Twelve Data account and obtain an API key to access their service.
- The API allows you to specify the stocks you're interested in and the desired data points (e.g., opening price, closing price, volume).

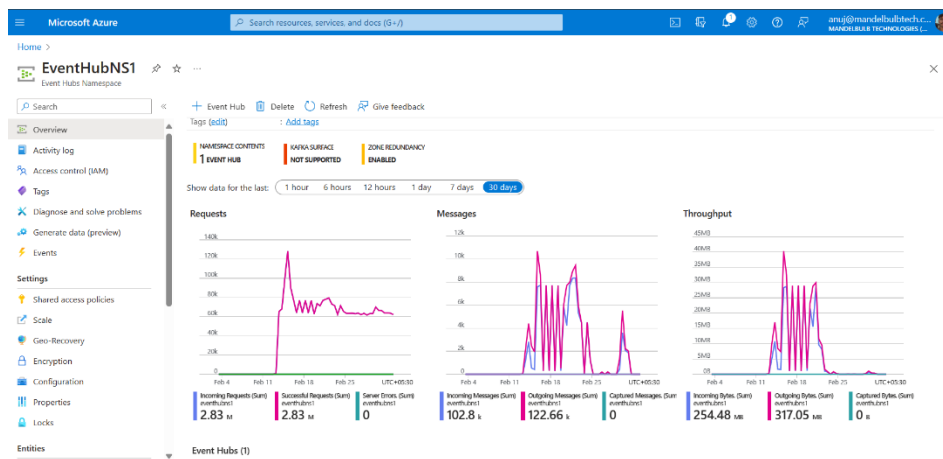
1.2.Data Fetching: Azure Logic Apps

- Azure Logic Apps is a cloud-based service for automating workflows by connecting to various Azure services and third-party APIs.
- In this phase, you'll create a Logic App that:
 - Triggers at regular intervals (e.g., every hour) using a built-in scheduler.
 - Uses an HTTP connector action to call the Twelve Data API with the desired stock information.
 - Parses the response from the API to extract the relevant stock price data.



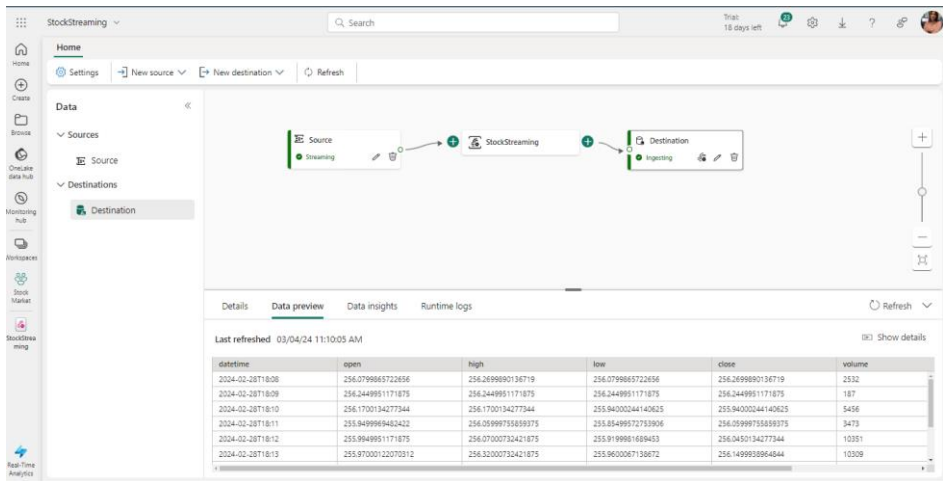
1.3.Data Streaming: Event Hub (Azure)

- Event Hub is a highly scalable real-time data ingestion service in Azure.
- It acts as a buffer, decoupling data producers (Logic Apps) from data consumers (KQL database) and handling high volumes of data streams.
- The Logic App will send the fetched stock price data as events to the configured Event Hub in your Azure subscription.



1.4. Data Streaming with Eventstream

- Data Source: Fetching Data from twelvedata API using logic apps
- Event Hub: This score is likely published to the Azure Event Hub using libraries like azure-eventhub-py. The Event Hub acts as a buffer, decoupling the data producer (Python notebook) from the data consumer (Eventstream).
- Eventstream: An Eventstream within Microsoft Fabric is configured to consume data from the Azure Event Hub.
- Event Transformation: (Optional) As mentioned earlier, you can utilize Eventstream's transformation capabilities to process the data further before sending it to KQL.
- Event Destination: The Eventstream is configured with a destination linked to the KQL database. This specifies the target KQL table where the combined sentiment scores will be stored.



1.5. Data Storage: KQL Database (Microsoft Fabric)

- KQL Database, powered by Kusto Query Language (KQL), is a service optimized for storing, analyzing, and visualizing large volumes of time-series data.
- The streamed stock price data from the Event Hub is ingested into the KQL database.
- You can use KQL to efficiently query, analyze, and visualize the historical stock price data for further insights.

Explore your data

Run Save as KQL queryset Copy query Build PowerBI report

1 NewsTable
2 take 1000

Table 1 Stats Search Done (0.123 s) 1,000 records

	Stock Symbol	FinBERT Score	RoBERTa Score	ChatGPT Score
www.ndbiprofit.com/business/coffee-craze-at-home...	NFLX	0.464432384239543	-0.004542231559753418	0.0
www.ndbiprofit.com/business/coffee-craze-at-home...	DIS	0.464432384239543	-0.004542231559753418	0.0
www.ndbiprofit.com/business/coffee-craze-at-home...	KO	0.464432384239543	-0.004542231559753418	0.1
www.ndbiprofit.com/business/coffee-craze-at-home...	AMZN	0.464432384239543	-0.004542231559753418	0.0
www.ndbiprofit.com/business/coffee-craze-at-home...	META	0.464432384239543	-0.004542231559753418	0.0
www.ndbiprofit.com/business/openal-says-it-catego...	CSCO	-0.7215047031641006	0.00791568540295166	-0.4
www.ndbiprofit.com/business/openal-says-it-catego...	ADSK	-0.7215047031641006	0.00791568540295166	-0.6
www.ndbiprofit.com/business/openal-says-it-catego...	INTC	-0.7215047031641006	0.00791568540295166	-0.6
www.ndbiprofit.com/business/openal-says-it-catego...	ADBE	-0.7215047031641006	0.00791568540295166	0.0
www.ndbiprofit.com/business/openal-says-it-catego...	WMT	-0.7215047031641006	0.00791568540295166	0.0
www.ndbiprofit.com/business/openal-says-it-catego...	NFLX	-0.7215047031641006	0.00791568540295166	-0.6
www.ndbiprofit.com/business/openal-says-it-catego...	DIS	-0.7215047031641006	0.00791568540295166	0.0

Explore your data

Run Save as KQL queryset Copy query Build PowerBI report

1 NewsTable
2 take 1000

Table 1 Stats Search Done (0.123 s) 1,000 records

Headline	Description	Source	Published
Coffee Craze Forces Asia's Big Exporters To Buy Brazilian Be...	Some of Asia's biggest coffee-producing nations are findin...	bloombergquint.com	2024
Coffee Craze Forces Asia's Big Exporters To Buy Brazilian Be...	Some of Asia's biggest coffee-producing nations are findin...	bloombergquint.com	2024
Coffee Craze Forces Asia's Big Exporters To Buy Brazilian Be...	Some of Asia's biggest coffee-producing nations are findin...	bloombergquint.com	2024
Coffee Craze Forces Asia's Big Exporters To Buy Brazilian Be...	Some of Asia's biggest coffee-producing nations are findin...	bloombergquint.com	2024
Coffee Craze Forces Asia's Big Exporters To Buy Brazilian Be...	Some of Asia's biggest coffee-producing nations are findin...	bloombergquint.com	2024
Coffee Craze Forces Asia's Big Exporters To Buy Brazilian Be...	Some of Asia's biggest coffee-producing nations are findin...	bloombergquint.com	2024
Coffee Craze Forces Asia's Big Exporters To Buy Brazilian Be...	Some of Asia's biggest coffee-producing nations are findin...	bloombergquint.com	2024
Coffee Craze Forces Asia's Big Exporters To Buy Brazilian Be...	Some of Asia's biggest coffee-producing nations are findin...	bloombergquint.com	2024
Coffee Craze Forces Asia's Big Exporters To Buy Brazilian Be...	Some of Asia's biggest coffee-producing nations are findin...	bloombergquint.com	2024
OpenAI Says It 'Categorically Disagrees' With Musk's Suit	OpenAI 'categorically disagrees' with the lawsuit Elon Musk...	bloombergquint.com	2024
OpenAI Says It 'Categorically Disagrees' With Musk's Suit	OpenAI 'categorically disagrees' with the lawsuit Elon Musk...	bloombergquint.com	2024

Phase 2: News Sentiment Analysis and Integration

2.1. Data Source: MarketAux API ([<https://marketaux.com>])

- MarketAux provides an API for accessing real-time and historical market news.
- Similar to Twelve Data, you'll need a MarketAux API key to retrieve relevant news articles.
- You can configure the API to filter news based on keywords related to the 10 US stocks you're tracking.

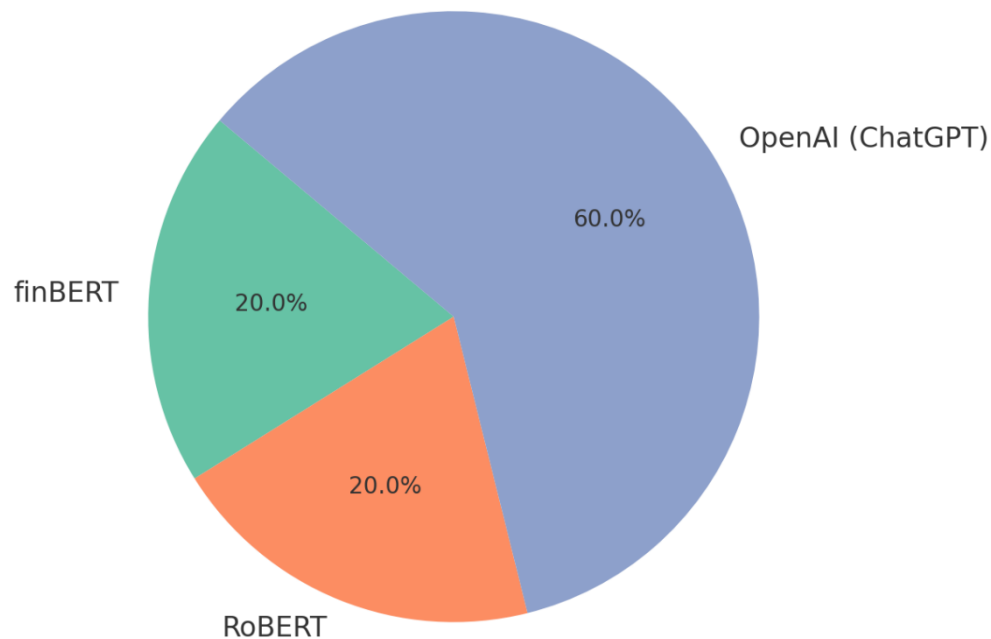
2.2. Data Processing: Python notebook in Microsoft Fabric

- Microsoft Fabric is a platform for developing and deploying dataflows.
- In this project, you'll create a Python notebook that runs within the Fabric environment.
- The notebook will:
 - Use Python libraries to interact with the MarketAux API.
 - Fetch news articles relevant to the tracked stocks based on the configured keywords.

2.3. Sentiment Analysis Tools:

- Sentiment analysis involves determining the emotional tone (positive, negative, or neutral) of a text piece.
 - This project utilizes three different tools for a more comprehensive analysis:
 - finBERT: This pre-trained model is specifically designed for understanding financial text, potentially offering better accuracy for news related to stocks.
 - RoBERT: A robust general-purpose model trained on a massive dataset of text and code, offering a broader understanding of language.
 - OpenAI (ChatGPT): A large language model known for its ability to interpret complex language nuances, potentially providing additional insights.
- #### 2.4. Score Combination:
- Each sentiment analysis tool assigns a score (positive, negative, or neutral) to each news article.
 - This project employs a weighted average to combine the scores from different tools:
 - finBERT: 20% weight - Capitalizes on its financial text expertise.
 - RoBERT: 20% weight - Leverages its general understanding of language.
 - OpenAI: 60% weight - Places greater emphasis on its ability to capture subtle nuances.

Weight Distribution for Sentiment Analysis Tools



2.5. Data Sending: Custom App in Microsoft Fabric

- A custom app developed within Microsoft Fabric is responsible for sending the combined sentiment score for each news article.
- This custom app likely utilizes Python libraries to format and send the data efficiently.

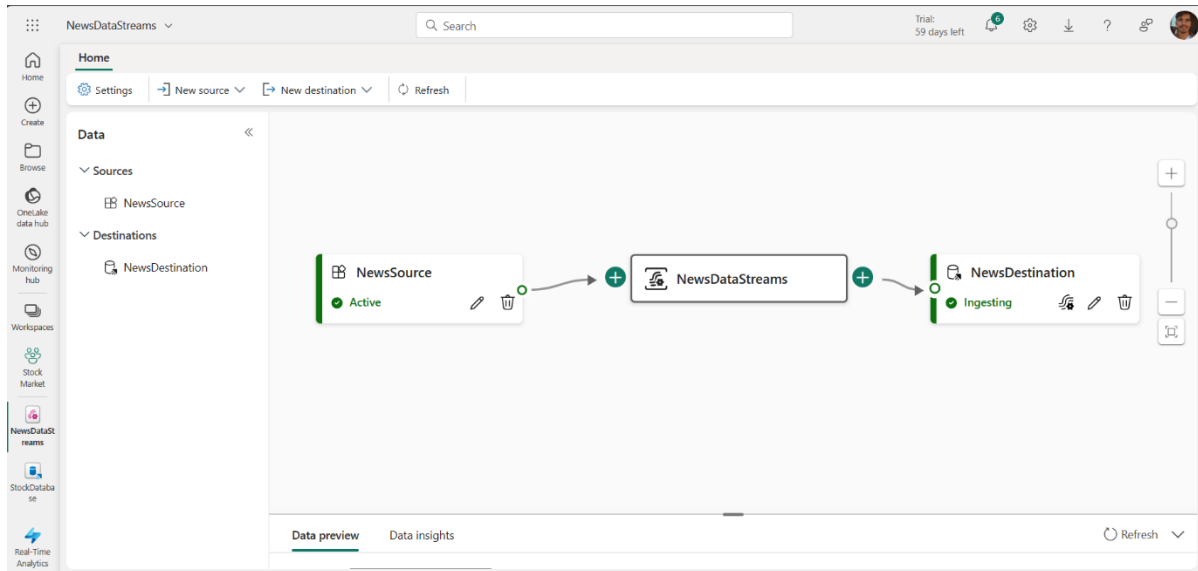
2.6. Data Streaming with Eventstream:

Eventstream is a component within Microsoft Fabric specifically designed for real-time data streaming and transformation.

In this project, an Eventstream is used to stream the combined sentiment score data.

- Event Source: The Python notebook acting as the data source generates the combined sentiment score for each news article.
- Event Transformation: (Optional) You can utilize Eventstream's built-in transformation capabilities within the event processor node to further refine the data before sending it to KQL. For example, you could add timestamps or format the data in a specific way.

- **Event Destination:** The final step involves configuring an Eventstream destination linked to the KQL database. This destination specifies the target KQL table where the data will be stored.



2.7. Data Storage: KQL Database (Microsoft Fabric)

- The combined sentiment scores, along with timestamps, are streamed into the KQL database alongside the corresponding stock price data.
- This allows you to analyze and potentially identify correlations between news sentiment and stock price movements.

Explore your data

Run Save as KQL queryset Copy query Build PowerBI report

```

1 PriceTable01
2 | order by [datetime]
3 | take 1000

```

Table 1 Stats Done (0.138 s) 1,000 records

	datetime	open	high	low	close	volume	symbol
>	2024-03-01T20:59	502.4800109853281	502.6499938964844	502.1700134277344	502.5400085449219	228156	META
>	2024-03-01T20:59	619.5800170898438	619.8800048828125	619.1300048828125	619.3400268554688	89125	NFLX
>	2024-03-01T20:59	48.3849983215332	48.41999816894531	48.380001068115234	48.38999938964844	454794	CSCO
>	2024-03-01T20:59	111.97000122070312	112.05000305175781	111.94000244140625	111.94000244140625	194850	DIS
>	2024-03-01T20:59	264.739990234375	264.9100036621094	264.6499938964844	264.67999267578125	49445	ADSK
>	2024-03-01T20:59	43.82500076293945	43.849998474121094	43.810001373291016	43.84000015258789	961694	INTC
>	2024-03-01T20:59	571.1644897460938	571.2000122070312	570.72998046875	570.739990234375	70048	ADBE
>	2024-03-01T20:59	59.564998626708984	59.56999969482422	59.52000045776367	59.529998779296875	432322	KD
>	2024-03-01T20:59	58.79499816894531	58.79999923706055	58.72999954223633	58.7400016784668	548716	WMT
>	2024-03-01T20:59	178.30499267578125	178.33999633789062	178.1699981689453	178.28999450683594	890617	AMZN
>	2024-03-01T20:58	58.76499938964844	58.79999923706055	58.75	58.790000915527344	179882	WMT
>	2024-03-01T20:58	48.375	48.3899938964844	48.35499954223633	48.3899938964844	184748	CSCO
>	2024-03-01T20:58	178.3480987548828	178.35000610351562	178.27999877929688	178.30499267578125	299604	AMZN

Further Analysis and Applications

- The combined sentiment data and stock prices can be used for various analyses, including:
- Exploring potential correlations between news sentiment and stock price movements.
- Identifying news events that may have a significant impact on specific stocks.
- Building machine learning models to predict stock price trends based on sentiment analysis.

Tools and Technologies

- Twelve Data API
- Azure Logic Apps
- Azure Event Hub
- KQL Database
- Microsoft Fabric
- Python
- finBERT
- RoBERT
- OpenAI (ChatGPT)