Bing News API Project Breakdown

This document provides a comprehensive overview of a large-scale Bing News API project, detailing the end-to-end data pipeline. It covers the various components involved in the project, including data ingestion, processing, transformation, orchestration, analytics, optimization, and the final business impact.

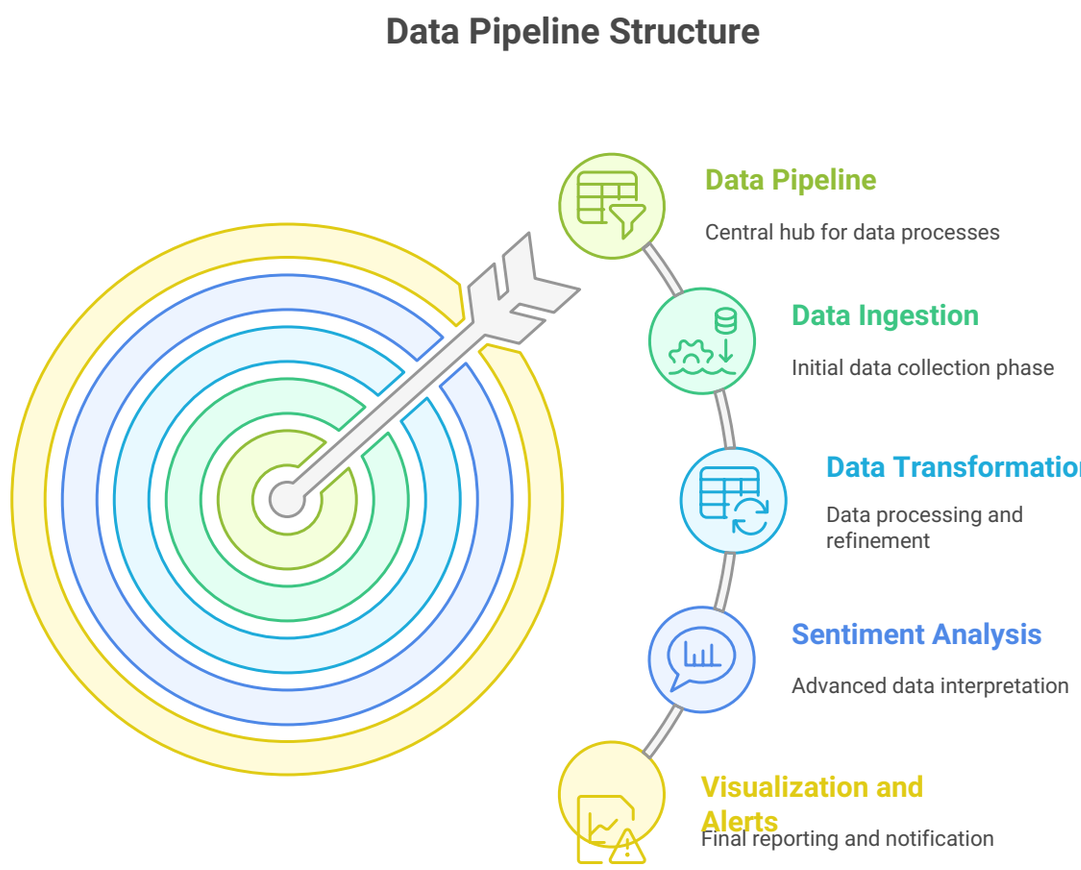
The project utilizes Microsoft Fabric to create a robust news analytics platform that delivers actionable insights into Bing News data.

Project Overview

The project focuses on ingesting, processing, and visualizing Bing News data, incorporating sentiment analysis and real-time alerts. The result is a powerful analytics platform that provides stakeholders with critical insights into the latest news trends.

Project Structure

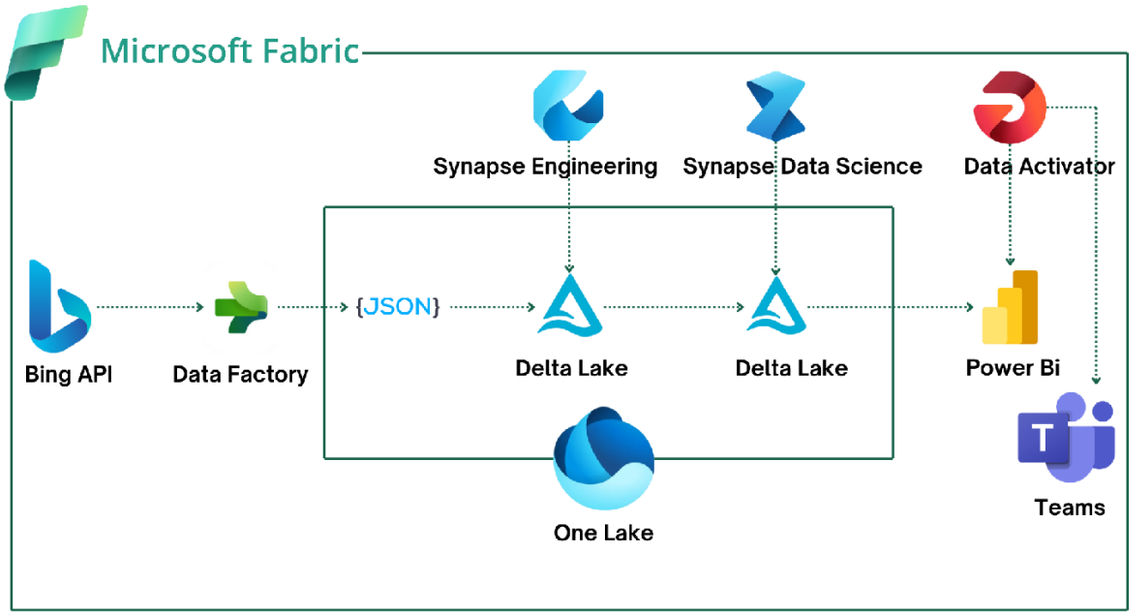
```
├── DataFactory (Data ingestion pipelines)
├── SynapseDataEngineering (Transformation)
├── SynapseDataScience (Sentiment Analysis) -- Incremental Load
├── PowerBI (Dashboard Reports)
├── DataActivator (Alert Configurations)
```



Tools and Technologies

- **Microsoft Fabric**: Unified platform for data engineering and data science
- **Data Factory**: Data ingestion pipelines

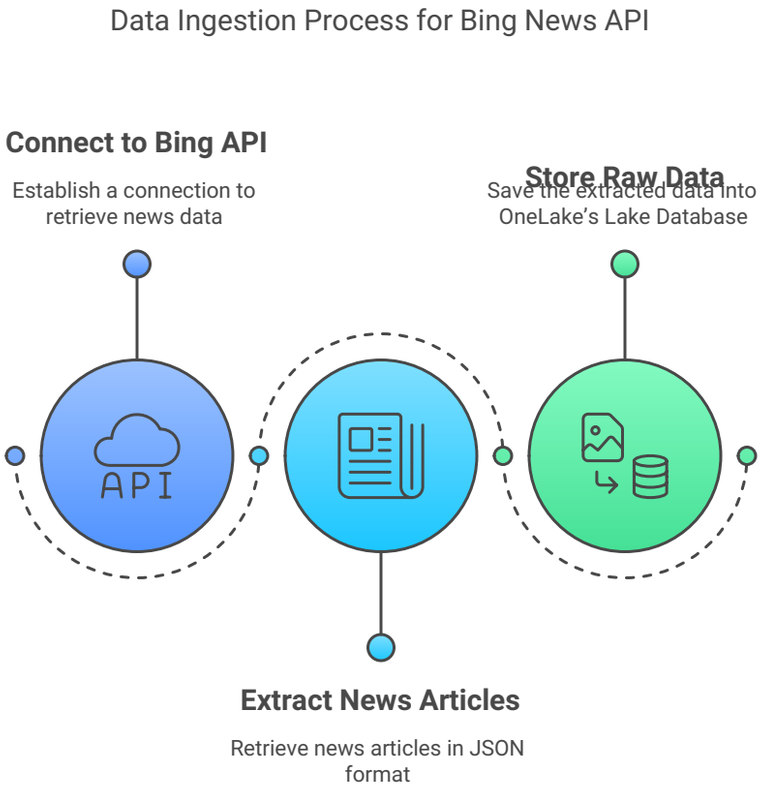
- **OneLake**: Centralized storage for raw and processed data
- **Synapse Data Engineering**: Data transformation with Spark Notebooks
- **Synapse Data Science**: Pre-trained machine learning models for sentiment analysis
- **Power BI**: Data visualization
- **Data Activator**: Real-time alert configurations



Key Features

1. Data Ingestion

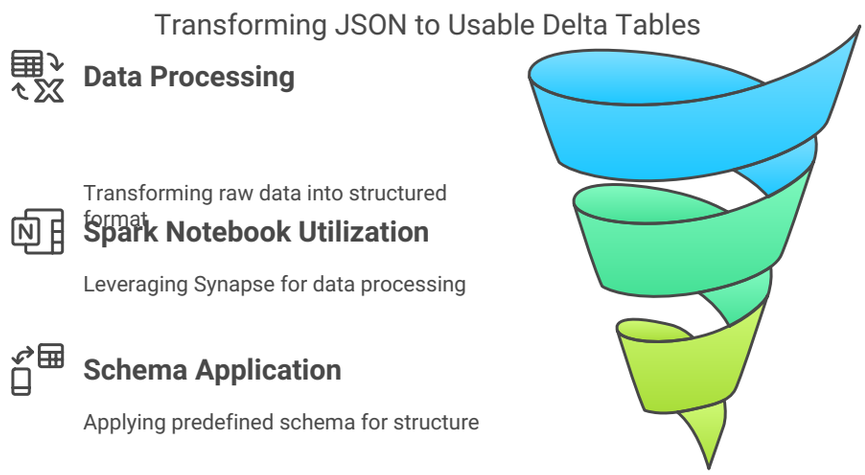
- Utilized Data Factory to connect to the Bing API.
- Extracted news articles in JSON format.
- Stored raw data in OneLake's Lake Database.



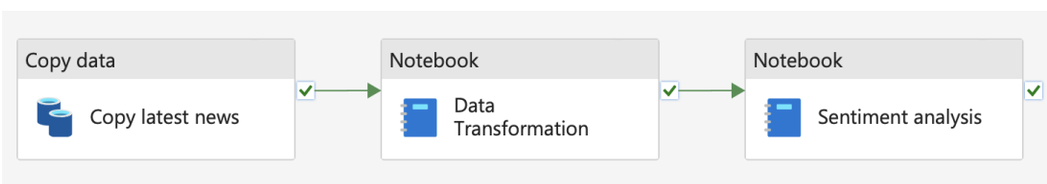
2. Data Transformation

- Processed raw JSON files into structured Delta tables.
- Utilized Spark Notebooks in Synapse Data Engineering.

- Applied a predefined schema for better usability.

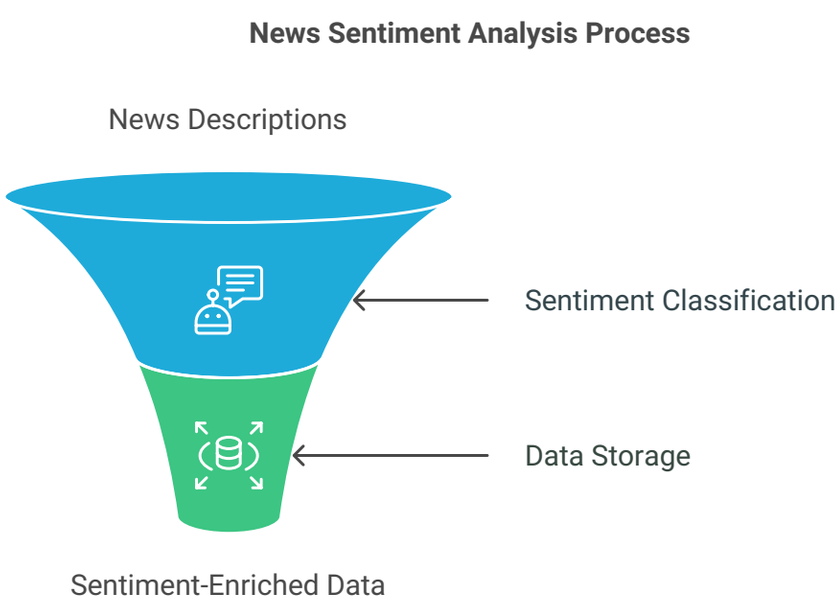


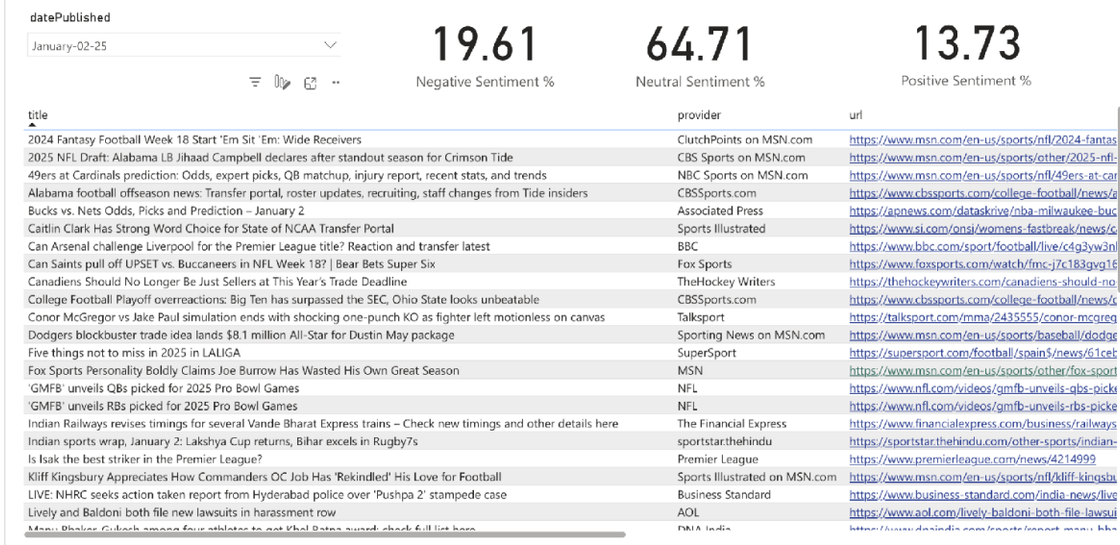
2.1 Data Pipeline



3. Sentiment Analysis

- Leveraged Synapse Data Science to analyze news descriptions.
- Used a pre-trained text analysis model to classify sentiment as -
    - Positive
    - Negative
    - Neutral
- Stored sentiment-enriched data as Delta tables in the Lake Database.





datePublished

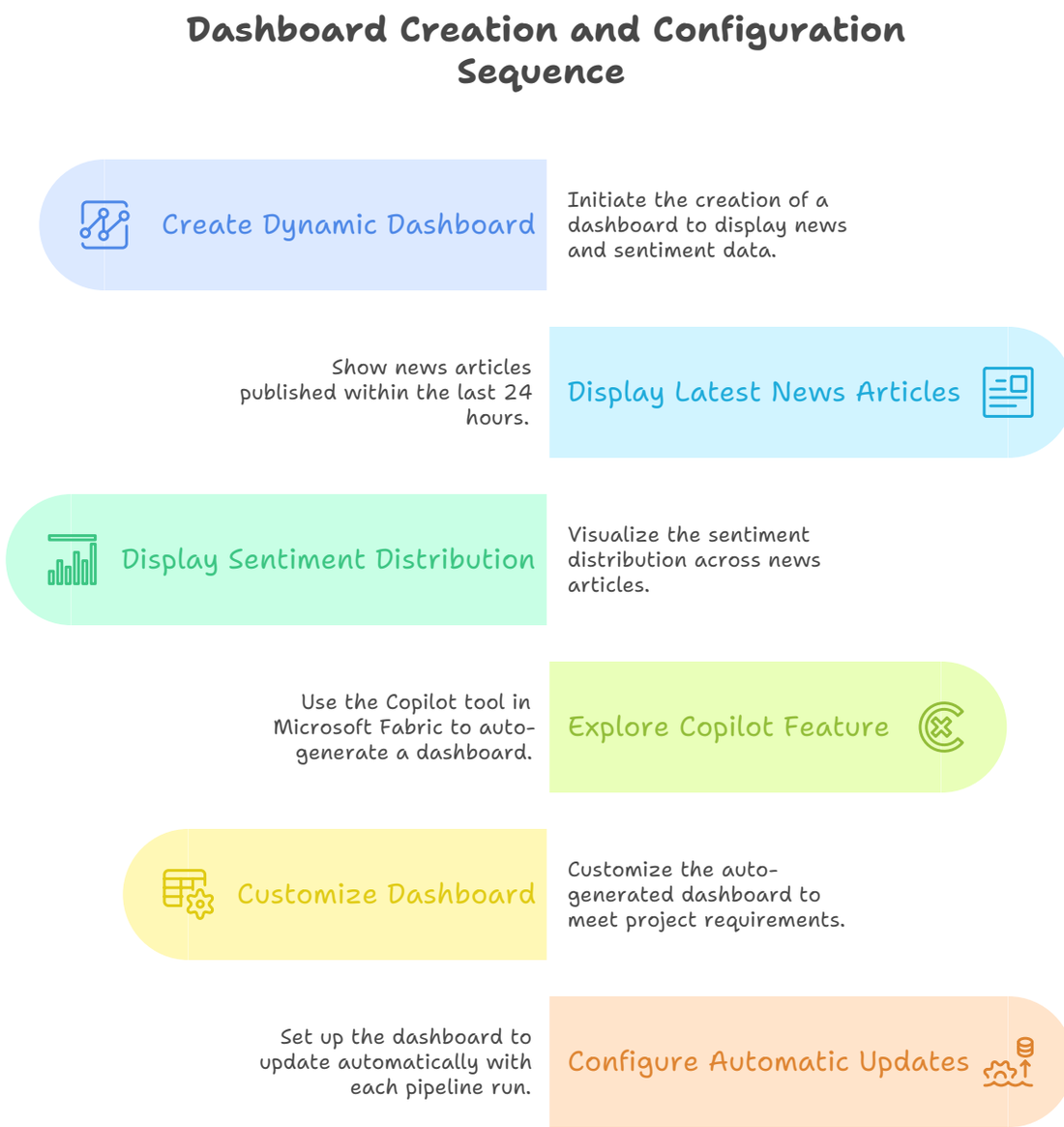
January-02-25

19.61	64.71	13.73
Negative Sentiment %	Neutral Sentiment %	Positive Sentiment %

title	provider	url
2024 Fantasy Football Week 18 Start 'Em Sit 'Em: Wide Receivers	ClutchPoints on MSN.com	https://www.msn.com/en-us/sports/nfl/2024-fantas
2025 NFL Draft: Alabama LB Jihaad Campbell declares after standout season for Crimson Tide	CBS Sports on MSN.com	https://www.msn.com/en-us/sports/other/2025-nfl-
49ers at Cardinals prediction: Odds, expert picks, QB matchup, injury report, recent stats, and trends	NBC Sports on MSN.com	https://www.msn.com/en-us/sports/nfl/49ers-at-car
Alabama football offseason news: Transfer portal, roster updates, recruiting, staff changes from Tide insiders	CBSSports.com	https://www.cbssports.com/college-football/news/a
Bucks vs. Nets Odds, Picks and Prediction – January 2	Associated Press	https://apnews.com/dataskrive/nba-milwaukee-buc
Caitlin Clark Has Strong Word Choice for State of NCAA Transfer Portal	Sports Illustrated	https://www.si.com/onsi/womens-fastbreak/news/c
Can Arsenal challenge Liverpool for the Premier League title? Reaction and transfer latest	BBC	https://www.bbc.com/sport/football/live/c4g3yw3nl
Can Saints pull off UPSET vs. Buccaneers in NFL Week 18? | Bear Bets Super Six	Fox Sports	https://www.foxsports.com/watch/fmc-j7c183gvg16
Canadiens Should No Longer Be Just Sellers at This Year's Trade Deadline	TheHockey Writers	https://thehockeywriters.com/canadiens-should-no
College Football Playoff overreactions: Big Ten has surpassed the SEC, Ohio State looks unbeatable	CBSSports.com	https://www.cbssports.com/college-football/news/c
Conor McGregor vs Jake Paul simulation ends with shocking one-punch KO as fighter left motionless on canvas	Talksport	https://talksport.com/mma/2435555/conor-mcgreg
Dodgers blockbuster trade idea lands $8.1 million All-Star for Dustin May package	Sporting News on MSN.com	https://www.msn.com/en-us/sports/baseball/dodge
Five things not to miss in 2025 in LALIGA	SuperSport	https://supersport.com/football/spain5/news/61ceb
Fox Sports Personality Boldly Claims Joe Burrow Has Wasted His Own Great Season	MSN	https://www.msn.com/en-us/sports/other/fox-sport
'GMFB' unveils QBs picked for 2025 Pro Bowl Games	NFL	https://www.nfl.com/videos/gmfb-unveils-qbs-picke
'GMFB' unveils RBs picked for 2025 Pro Bowl Games	NFL	https://www.nfl.com/videos/gmfb-unveils-rbs-picke
Indian Railways revises timings for several Vande Bharat Express trains – Check new timings and other details here	The Financial Express	https://www.financialexpress.com/business/railways
Indian sports wrap, January 2: Lakshya Cup returns, Bihar excels in Rugby7s	sportstar.thehindu	https://sportstar.thehindu.com/other-sports/indian-
Is Isak the best striker in the Premier League?	Premier League	https://www.premierleague.com/news/4214999
Kliff Kingsbury Appreciates How Commanders OC Job Has 'Rekindled' His Love for Football	Sports Illustrated on MSN.com	https://www.msn.com/en-us/sports/nfl/kliff-kingsbu
LIVE: NHRC seeks action taken report from Hyderabad police over 'Pushpa 2' stampede case	Business Standard	https://www.business-standard.com/india-news/live
Lively and Baldoni both file new lawsuits in harassment row	AOL	https://www.aol.com/lively-baldoni-both-file-lawsui
Manu Bhaker, Gukesh among four athletes to get Khel Ratna award; check full list here	DNA India	https://www.dnaindia.com/sports/report-manu-bha

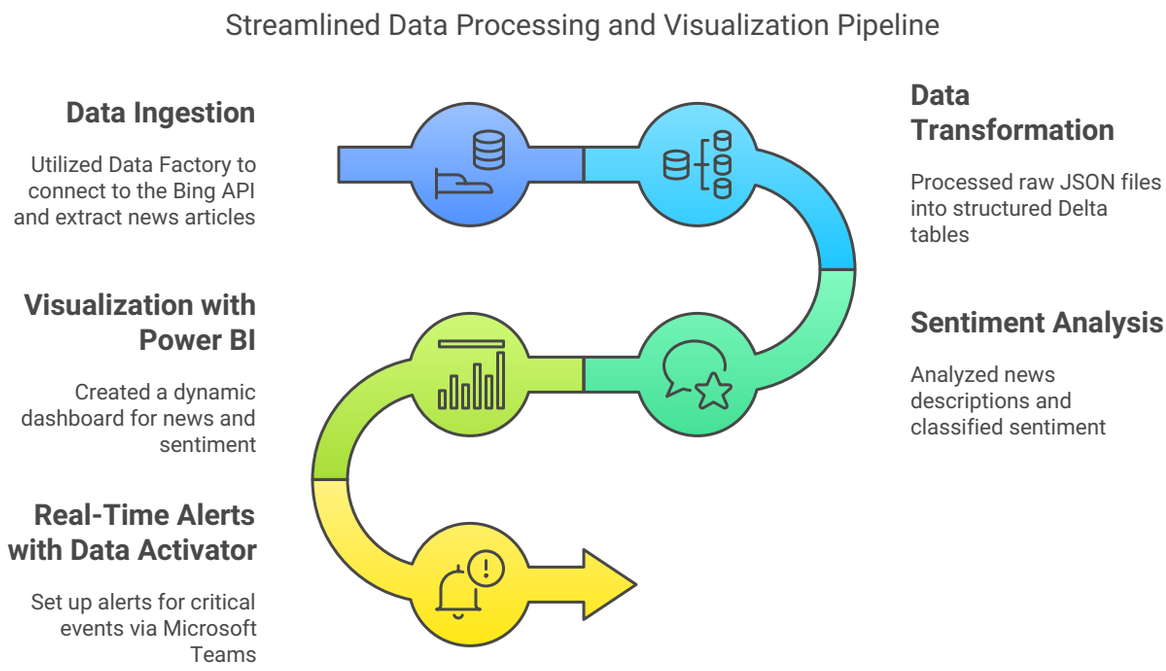
4. Visualization with Power BI

- Created a dynamic dashboard to display -
  - Latest news articles (published within the last 24 hours).
  - Sentiment distribution across news articles.
- Explored Copilot Feature: Leveraged the Copilot tool in Microsoft Fabric to auto-generate a dashboard, which was then customized to meet project requirements.
- Configured to update automatically with each pipeline run.



5. Real-Time Alerts with Data Activator

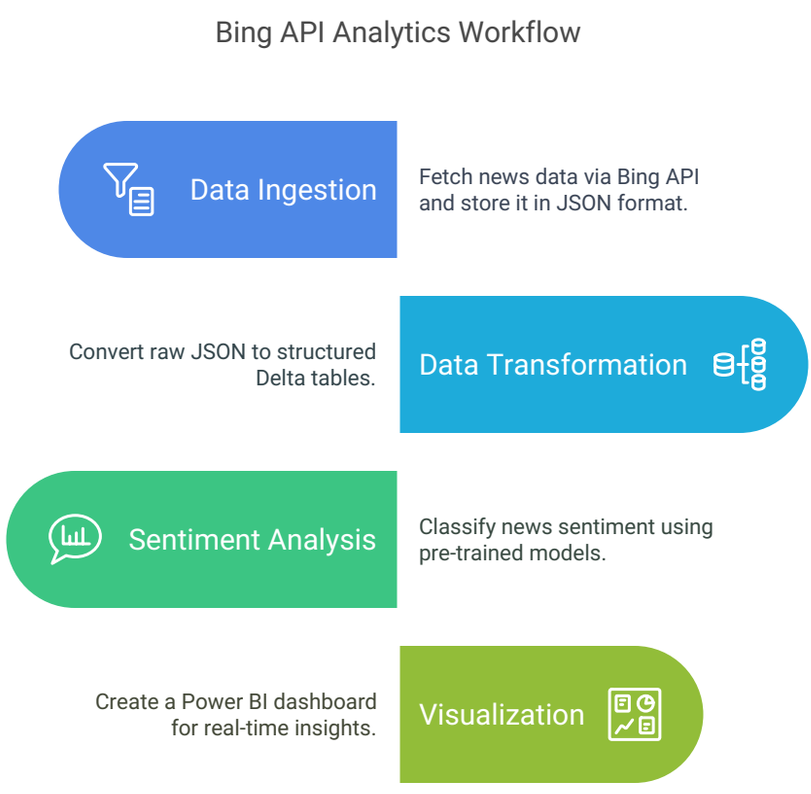
- Set up alerts for critical events (e.g., negative sentiment articles).
- Notifications delivered via Microsoft Teams.



Workflow

Step-by-Step Process:

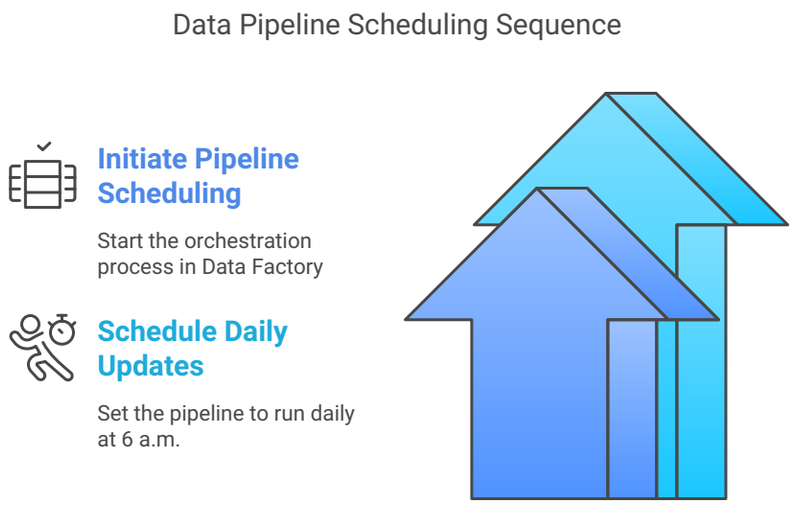
1. **Data Ingestion**: Fetch news data via Bing API and store it in JSON format.
2. **Data Transformation**: Convert raw JSON to structured Delta tables.
3. **Sentiment Analysis**: Classify news sentiment using pre-trained models.
4. **Visualization**: Create a Power BI dashboard for real-time insights.
5. **Alerting**: Configure alerts for significant updates using Data Activator.



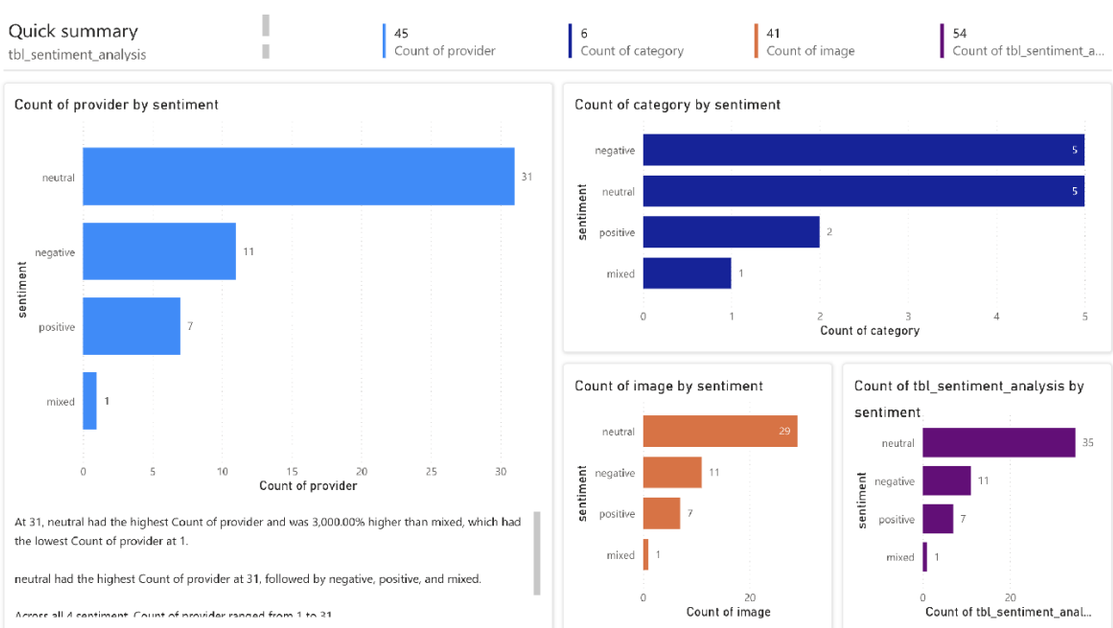
Alerting Configure alerts for significant updates using Data Activator.

Automation

- **Pipeline Scheduling**: Orchestrated the entire workflow in Data Factory.
- **Daily Updates**: Scheduled pipeline to run at 6 a.m. every day.

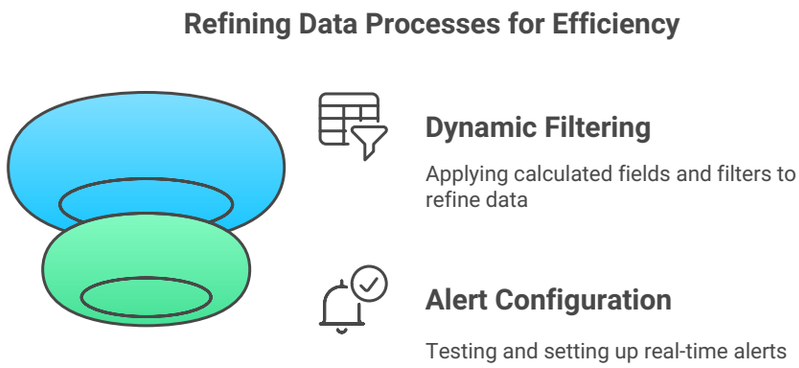


Power BI Dashboard



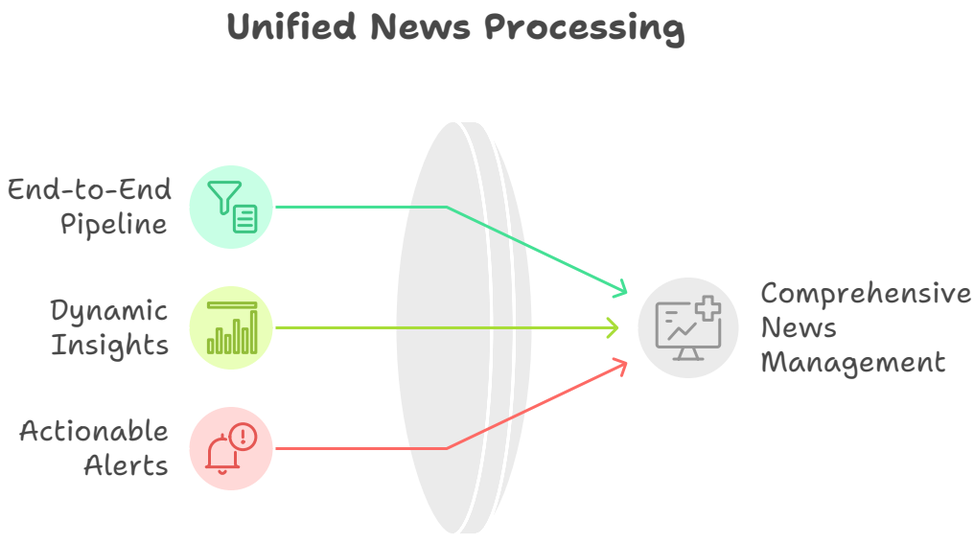
Challenges

- **Dynamic Filtering**: Initially challenging to display news from the last 24 hours; resolved using calculated fields and advanced Power BI filters.
- **Alert Configuration**: Setting up real-time alerts in Data Activator required rigorous testing; successfully implemented Microsoft Teams notifications.



Results

- **End-to-End Pipeline**: Fully automated and scalable.
- **Dynamic Insights**: Real-time dashboard to monitor news.
- **Actionable Alerts**: Instant notifications for critical events.



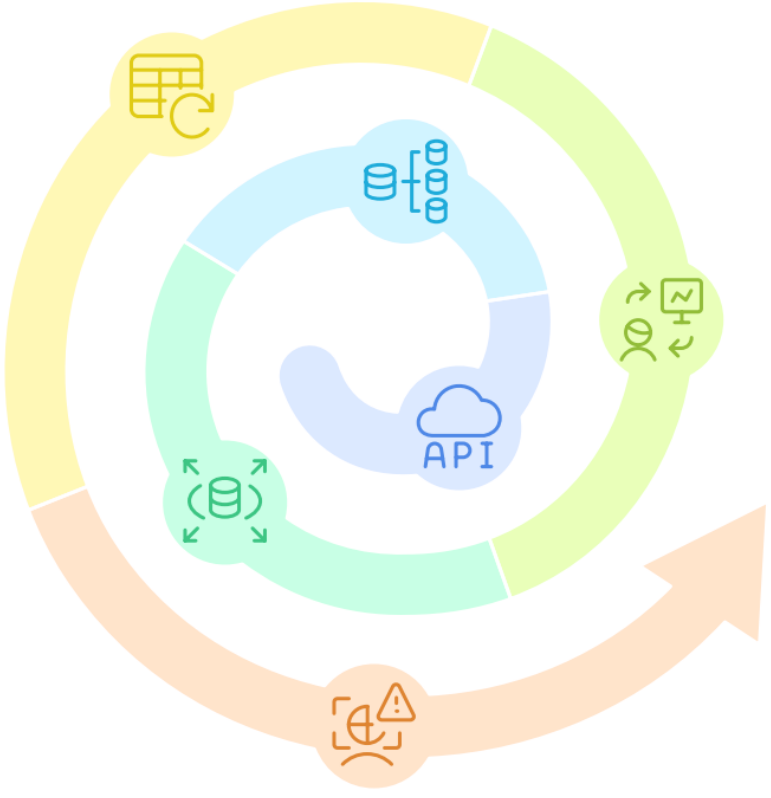
Key Files

- **DataFactory**: Bing API pipeline configurations.
- **Synapse Notebooks**: Data transformation and sentiment analysis scripts.
- **Power BI Dashboard**: Visualizations and filters.
- **Alert Configuration**: Data Activator setup for real-time notifications.

How to Run

1. **Setup Bing API**:
    - Configure API key in Azure.
    - Update credentials in Data Factory pipeline.
2. **Run Data Pipeline**:
    - Execute Data Factory pipeline for data ingestion.
3. **Execute Spark Notebooks**:
    - Transform JSON to Delta tables using Synapse Data Engineering.
4. **Perform Sentiment Analysis**:
    - Run Synapse Data Science scripts for sentiment classification.
5. **Visualize in Power BI**:
    - Open Power BI report and refresh to view the latest data.
6. **Test Alerts**:
    - Trigger alerts by adding negative sentiment articles.

Bing API Analytics Workflow



01 Setup Bing API

02 Run Data Pipeline

03 Execute Spark Notebooks

04 Perform Sentiment Analysis

05 Visualize in Power BI

06 Test Alerts

Access the files here - https://github.com/Soniaranvir/End-to-End-Data-Engineer