**Fabric Data Project**

A diagram of data analysis

AI-generated content may be incorrect.

## **1. Project Overview**

This project demonstrates a complete data engineering workflow using **Microsoft Fabric** by ingesting real-time data from the **Bing News Search API**, transforming and applying **Sentiment detection using** **Machine Learning** with **Pyspark Notebooks**, and visualizing insights using **Power BI**. A pipeline was also developed to **schedule daily runs** and send **real-time alerts** for specific triggers.

## **2. Workspace and Lakehouse Setup**

* Created a new Microsoft Fabric **Workspace**
* Created a **Lakehouse** to store both raw files and transformed tables

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

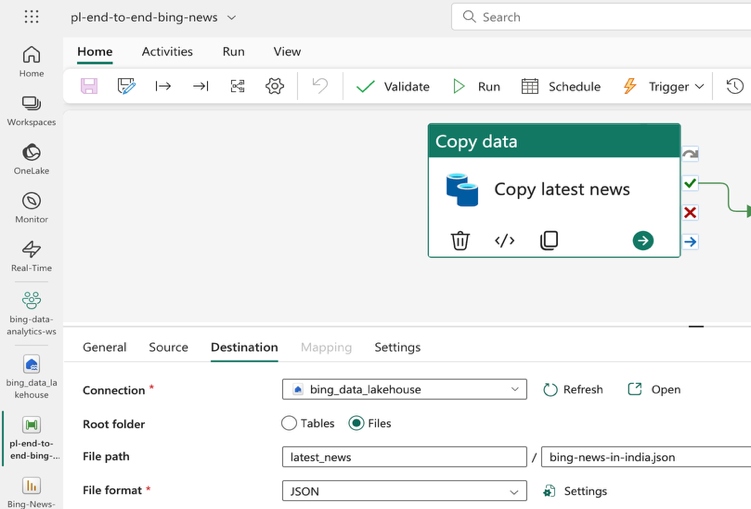
AI-generated content may be incorrect.

## **3. Data Ingestion (Bronze Layer)**

### 3.1 Pipeline Creation

* Created a new **Data Factory pipeline**
* Used the **REST API connector** to connect to Bing News API
* Added required headers (e.g., Ocp-Apim-Subscription-Key)
* Configured parameters like:
* q=latest+news&count=100&freshness=Day&mkt=en-IN

A screenshot of a computer

AI-generated content may be incorrect.   
 **3.2 Saving Raw Data**

* Target location: Lakehouse > Files section
* Saved data in **JSON** format
* Verified pipeline run status and inspected saved files

A screenshot of a computer

AI-generated content may be incorrect.

## **4. 🧹 Data Transformation (Silver Layer)**

* Created a new **PySpark Notebook**
* Connected the Lakehouse and loaded JSON data
* Performed Data Cleaning and Data Transformations like flattening,
* Incrementally Loaded the transformed data into a Delta Table named today\_news\_in\_India

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

## **5. 📊 Sentiment Analysis (Gold Layer)**

* Created another notebook for **Machine Learning-based sentiment detection analysis**
* Loaded the cleaned data table today\_news\_in\_India
* Clasified the articles as Positive, Negative, or Neutral
* Saved the results as a new Delta Table named news\_sentiment\_analysis

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

## **6. 📈 Power BI Reporting**

* Created a **Semantic Model** called daily\_news\_dashboard on top of the gold layer
* Built a Power BI report using:
  + News Title, Sentiment, Source, Publish Time
  + Filters like date and keyword search

### A screenshot of a computer AI-generated content may be incorrect.

### A screenshot of a news report AI-generated content may be incorrect.

### A screenshot of a computer AI-generated content may be incorrect.

## **6.1 Scheduled Refresh**

* Turned off Direct Lake refresh from the semantic model settings
* Refreshing the semantic model from Pipeline

A screenshot of a computer

AI-generated content may be incorrect.

## **7. 🔁 Pipeline Orchestration**

* Connected ingestion, transformation, and ML notebooks in a **single orchestrated pipeline**
* Scheduled to run **daily at 6 am for a period of time**

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

## **8. 🚨 Real-Time Alerts with Data Activator**

* Enabled alerts via **Power BI Data Activator**
* Added an alert on visual elements like the number of negative news items
* Configured alert actions and schedules

A screenshot of a computer

AI-generated content may be incorrect.

## **9. ✅ End-to-End Testing**

* Ran the full pipeline manually
* Verified successful refresh and updated visualizations

A screenshot of a calendar

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

## **10. 📂 Resources**

* Code Notebooks, Power BI Report File (PBIX) are available in : [Github-Link](https://github.com/Puneeth0106/Fabric_Projects/tree/master/Bing-News)

## **11. 🚀 Conclusion**

This project showcases how Microsoft Fabric can be used to design a full-scale data pipeline—from ingesting REST API data to transforming and analyzing it, to reporting and setting up actionable alerts. This solution is scalable, dynamic, and efficient for real-time data analytics.